DOCUMENT RESUME

ED 038 107 HE 001 485

TITLE	A Student Flow Information System for the State of California; A Report To The Joint Committee On Higher Education of the California State Legislature.
INSTITUTION	Aero-Jet General Corp., Sacramento, Calif.
PUB DATE	14 Feb 69
NOTE	100p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$5.10
DESCRIPTORS	*College Graduates, *Data Processing, *Educational Research, *Higher Education, *Information Systems
IDENTIFIERS	Aerojet General Corporation, California, *Service Bureau

ABSTRACT

This study proposes an information system to gather, coordinate, and evaluate longitudinal data collected periodically from students in sample groups beginning with entry into high school and continuing through young adulthood, i.e., five years post college. Its purpose would be to gain insight into characteristics of the student population, and thereby provide information for improving policies for higher education. The report provides background for the study and discusses the many aspects of establishing a contralized processing facility. The document contains tables, lists, and appendices. It is recommended that one responsible body, a service bureau, be established to coordinate information into a centralized data processing system. (NF)

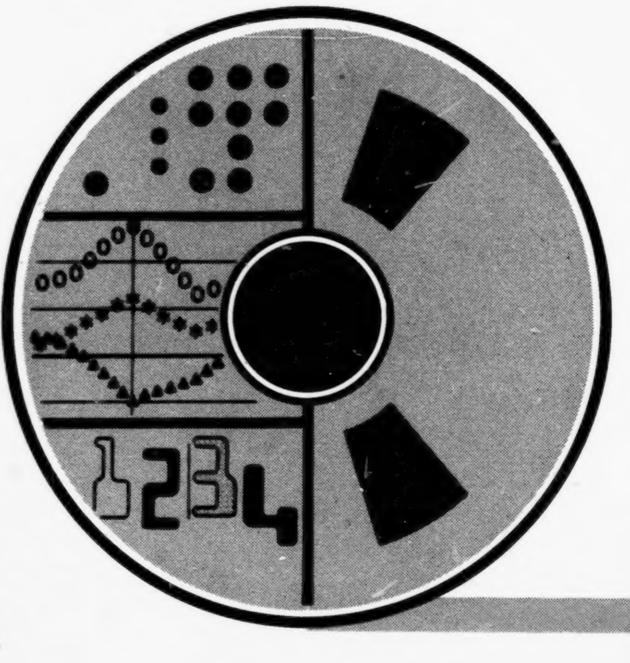
ED038107

A STUDENT FLOW INFORMATION SYSTEM
for the
STATE OF CALIFORNIA

A Report To

THE JOINT COMMITTEE ON HIGHER EDUCATION
of the
CALIFORNIA STATE LEGISLATURE

14 February 1969



U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

HE001 485

computing sciences

AEROJET-GENERAL CORPORATION

PREFACE

The California Joint Legislative Committee on Higher Education contracted, through Baxter McDonald and Company, with the Computing Sciences Division of Aerojet-General Corporation to prepare a conceptual design of a Student Flow Information System. The scope of the study was to set forth the design only in sufficient detail to provide the legislature with cost estimates and other information necessary to make a funding decision. The time for research, design and cost estimation encompassed a three month period beginning in October of 1968. Hence, this report, a culmination of that study, presents details only to the extent required to determine the feasibility and approximate cost of various alternatives.

The Aerojet-General Staff members, G. W. Fagerlin, Dr. T. N. Throckmorton, and A. C. Wells, wish to express their appreciation to the many educators and the personnel of SCOPE and Project TALENT who gave of their time and knowledge to enable the authors to compile this report. In addition, we would like to express our appreciation to J. E. Evans of the Staff of the Joint Committee and A. W. Baxter of Baxter, McDonald and Company, for their constructive comments and many stimulating discussions.

TABLE OF CONTENTS

Page

I. Introduction 1

II. Research and Analysis 6
  A. Review of Available Information 6
    1. Initial Study 6
    2. Interviews With Educational Institutions 9
    3. Review of TALENT and SCOPE Experience 13
  B. Decisions 18
    1. Selection of Data Elements 18
    2. How to Collect the Data 22
    3. The Role of the Institution in Data Collection 22
    4. Sample Size for Institutional Data Collection 25
    5. Followup Considerations 28

III. Systems Design 35
  A. Definition 35
  B. Data Collection 37
    1. Collecting Data From Students 41
    2. Collecting Data From Former Students 42
  C. Central Processing 44
  D. Implementation of System 48

IV. Costs 50
  A. Design and Programming 50
  B. Operation 51
    1. Educational Populations 51
    2. Data Collection 55
    3. Central Processing Costs 66
  C. Staffing Costs 68
  D. Cost Summary 72

TABLE OF CONTENTS (cont.)

Page

V. Summary 77

A. Coordination 77

B. Organizational Implications 79

C. Confidentiality 80

D. Performance 82

E. Timing 82

References 84

APPENDICES

A Preliminary Information Elements 85

B Personnel Interviewed 90

C Data Elements for Students While Enrolled in California Educational Institutions 92

D Final Data Elements 94

E Followup Data 98

F Incremental Characters 101

TABLE LIST

No.		Page
II-1	Fall 1967 Enrollment - California Schools	8
II-2	Necessity vs Effort	19
II-3	Expected Proportion, R, Remaining in a Sample After N Years of Followup	31
II-4	Estimated Sample Sizes Per Cell	32
II-5	Followup Population	34
IV-1	Projected Population Data	53
IV-2	Total Characters in Millions	59
IV-3	Incremental Data Collection Characters	59
IV-4	Incremental Characters in Millions	60
IV-5	Collection Costs	61
IV-6	Followup Costs	65
IV-7	Cost Summary	73

FIGURE LIST

No.		Page
II-1	Educational Institution Capability vs Time	25
III-1	Data Collection	38
III-2	Update Process	39
III-3	Report Generation Process	40
III-4	Followup Process	43
IV-1	System Population	57
IV-2	Collection Cost	62
IV-3	Central Processing Costs	69
IV-4	Project Staffing Costs	71
IV-5	Total Systems Cost by Year	74

I. INTRODUCTION

The California Joint Legislative Committee on Higher Education develops new legislation and prepares recommendations on other legislative proposals affecting the California system of higher education. In order to support these actions with substantive information, the Committee desires to gather factual data about students as they progress into, through, and out of the educational system. The data currently available through cross-sectional statistics only permit comparison of enrollments at a particular point in time with similar statistics gathered at a different point in time. There is little in these enrollment statistics that permits examination of the dynamics of the educational population.

In the opinion of the Committee, the information desired should be derived from longitudinal data collected periodically from individual students in sample groups beginning with their entry into the high schools and continuing through to young adulthood. Using such data, the Committee expects to determine the accessibility of higher education to students of different backgrounds and to gain better insight into the flow and persistence characteristics of the student population as a whole. In particular, the Committee is seeking indicators which, after further investigation, may lead to legislation that will offset the undesirable and preventable forms of student attrition and expand the accessibility of higher education to students of varying background and ability. Further, the Committee expects to be able to develop significantly improved policies and programs for the higher educational system and be able to measure the success of these at some future date.

To determine methods of acquiring, storing and processing the desired factual data, a contract was awarded to the Computing Sciences Division at the Sacramento facility of Aerojet-General Corporation, to prepare a conceptual design and estimated costs for the development of an appropriate system. The contractual agreement specified that the research design and cost estimation phases were to be completed within a period of three months. Aerojet was to determine the types of information to be collected, the samples to be used and the method and frequency of collecting data. The final report of Aerojet was to indicate the estimated costs associated with the development, final implementation and operation of a statewide system. On the basis of these several factors, the Committee will make recommendations to the legislature if further funding is warranted.

The Joint Committee is concerned that less than 60% of the students entering California High Schools graduate and enter higher educational institutions. Of equal concern is that only 20% of those entering high schools obtain a bachelor's degree. This information, estimated from cross-sectional studies, indicates a condition which may be good or bad, but certainly warrants further investigation. The additional investigation must provide answers to such questions as:

- Who benefits directly from our public system of higher education and who does not?
- What is the distribution and flow of students within the educational system?
- In what types of educational institutions does the student receive the greater benefit?
- What should be the functions of the different types of educational institutions?
- What state support should be allocated to the different types of institutions for operations, new facilities, new faculty positions, new academic departments, and the like?

I, Introduction (cont.)

- Should increasing financial support be afforded a system of education in which the output, measured in graduated students with bachelor's degrees, seems so meager?
- How many California High Schools and Junior College Graduates earn their degrees at out-of-state colleges and universities?
- Are the other-than-tuition costs prohibitive except to the more well-to-do families?
- Should the state provide greater financial assistance to lower income groups to increase the number of students obtaining a bachelor's degree?
- How should student aid be allocated and how much should students be assessed?
- Are the admissions policies of our educational institutions appropriate to the changing local, state and national environments?
- Do the California policies of higher education need revision?

These questions cannot be answered with enrollment data alone. On the contrary, the information required must be collected from the individual student and the institution. The information needed falls into several general categories as follows:

FAMILY BACKGROUND - Are the students' parents college graduates, low, high or middle income, rural or urban, etc.?

SOCIO-ECONOMIC - Is the student single or married, how many dependents does he have, is he working while in school, is his employment and enrollment full or part-time, etc.?

ABILITY AND ACHIEVEMENT RECORD - What is the student's grade point average and academic aptitude. What degrees has he already earned, etc.?

LENGTH OF TIME SPENT IN SCHOOL - Did he stay in school as a full-time student until completion, was his degree obtained in four, five or six years? Did he go to junior college for one or two years, etc.?

BENEFITS RECEIVED FROM EDUCATION - Is he employed full or part-time, is his job related to his education, what is his job level versus years of schooling, etc.?

The answers to the preceding questions, after being subjected to analysis, should provide the kinds and types of information needed for legislative evaluation of such problems as attrition, persistence and flow characteristics of the educational population. In addition, it is felt that some subjective evaluation of the benefits the individual derives from his educational experience could be obtained.

Knowing the types of information needed, an inventory of data and methodology was established. This inventory encompassed other known research projects as well as the registration and admissions operations of the educational institutions. In the former category were SCOPE and Project TALENT both of which are large-scale longitudinal studies with offices in California. Discussions with these groups provided some data on the mechanics of and problems associated with followup studies. Visits to some of the high schools, colleges, and university campuses determined that information which is normally collected during admissions and registration.

The results of the Aerojet work is reported in the following chapters. While the report is general, there are sufficient details upon which to judge the merits and cost of the proposed Student Flow Information System and thus whether more detailed design and analysis are warranted.

II. RESEARCH AND ANALYSIS

This chapter describes the approach followed by Aerojet-General in investigating the problem stated in Chapter I. It provides a summary of information obtained during the study and discusses the reasons for certain decisions basic to the design of the Student Flow Information System. These decisions relate to the selection of data elements and the general method of data collection.

A. REVIEW OF AVAILABLE INFORMATION

1. Initial Study

In order to insure a mutual understanding of the problem between the contractor and the Joint Committee and to obtain greater insight into the type of data required, Aerojet-General developed an initial list of information elements that might be collected from or about selected populations of students entering, progressing through, and leaving the California Educational System. This initial list was submitted to an extensive review in a joint meeting with Baxter, McDonald & Company and staff consultants to the Joint Committee on Higher Education.

Initial discussions during this meeting were directed at the need for data which would determine both the flow of students, into, through, and out of the secondary and higher educational institutions and the reasons why trends in these flows occur. Although the reasons for certain trends are of the utmost significance, it became apparent that sufficient data to be clearly definitive in this regard would be both voluminous and difficult to obtain. For this reason data selected for retention during this review was based primarily upon determining the characteristics and number of students which follow various paths. This data in itself should provide considerable insight into probable causes while providing a clear definition of trends. It would establish an ideal base from which specialized studies could be launched to concentrate on more specific causes for those trends. The list of information elements agreed upon during the review is included as Appendix A. This list was intended only as a preliminary list, subject to additions and eliminations during subsequent phases of the study.

After establishing the data list, efforts were directed toward determining the best method of data collection. In order to gain perspective concerning the scope and relative magnitude of the data collection requirements, the 1967 fall enrollments in California schools, both secondary and higher, public and private, were obtained from the Department of Finance (8). This data is displayed in Table II-1 showing a total enrollment of nearly 2.2 million students. Of this total approximately 58 percent were high school students. Nearly 57 percent of the college students were enrolled in public junior colleges while about 12 percent were in private colleges. Considering the size and distribution of this population over various types of institutions, along with the diverse characteristics of students and the complexity of movement taking place, it was apparent that the collection of longitudinal data would be a substantial task.

Two basic alternative methods were considered. The first was to obtain data from the educational institutions utilizing existing documents and established data collection processes such as registration. The second was to obtain data directly from the individuals through follow-up procedures similar to those employed by the Project TALENT and SCOPE. In order to evaluate which or what combination of these procedures provided the best data collection approach, a series of interviews were conducted, the results of which are presented below.

TABLE II-1

FALL 1967 ENROLLMENTS - CALIFORNIA SCHOOLS

		PUBLIC		PRIVATE		TOTAL
		%	TOTAL	%	TOTAL	
HIGH SCHOOLS	FR	28	1,184,806	30	84,847*	1,269,653
	SO	27		26		
	JR	24		23		
	SR	21		21		
JUNIOR COLLEGES	FR	71	521,695		112,146	914,809
	SO	24				
STATE COLLEGES			185,601			
UNIVERSITIES			95,367			
TOTAL			1,987,469		196,993	2,184,462

Fall 1967 Enrollment Data from Department of Finance
State of California

* Excludes non-sectarian Schools

2. Interviews with Educational Institutions

The list of information elements in Appendix A was reviewed during on-site interviews with representatives of school districts, junior colleges, state colleges and universities. Personnel and locations visited are listed in Appendix B. The locations visited were selected by the Aerojet-General staff based on convenience of access and an acquaintance with most of the personnel to be interviewed. In so doing, the contractor was able to acquire considerable information in the short time available for the visits. While these locations were not selected by any recognized sampling technique, it was felt they represented a reasonable cross section. Further, the representatives interviewed have gained knowledge of the operations at other campuses through years of participation in professional groups and conferences. The interviews were directed primarily at what data was currently collected, how it was collected, and in what form it was available. Similarly the problems associated with collecting certain specialized types of information such as ethnic group, occupation and salary of parents, etc. were discussed. The following paragraphs summarize the results of these discussions.

Forms used for collection of data from students are standardized only within the University of California System. Different applications forms are used for undergraduate and graduate students. Both the state colleges and the universities use card packets (punched card size) for registration. Each Junior College Districts visited collected registration data using different combinations of cards and printed forms. The University of the Pacific has an 8-1/2 x 5-1/2 registration booklet with tear-out pages. The elementary-secondary school districts use combinations of punched cards and printed forms of their own design. Transcripts of a student's previous academic record are usually received under separate cover at the time of application.

Admissions data from the application is collected for each new or re-entering student on a one-time basis at the colleges and universities. Registration data is generally collected at the beginning of each quarter or semester although some specialized data may be obtained annually. Enrollment data at the elementary-secondary level is collected at the beginning of each school year.

At all levels, the institutions collect from each student: name, address (permanent and/or current), sex, date of birth, and state or country of birth. Selective service numbers are collected on those students desiring reports to their draft boards. Although social security numbers are not required at all campuses, each obtains it if available. The University of California and University of the Pacific now require it of all students and most colleges are moving in that direction. Social security numbers are not normally collected at the high school level although some schools are beginning to use them.

The name and address of a parent or guardian were the only parental family data generally collected. Most institutions do not require this for students over 21. More complete data is collected on the parental family for those students requesting financial assistance, but is usually filed separately in financial aid offices. Aside from marital status, which is collected at all levels, marital family data was collected only on an exceptions basis.

At the secondary level, cumulative folders usually provide a history of elementary and high schools attended. Computer systems keep only the previous school attended regardless of whether it was an elementary or high school. For all college freshmen the institutions collect high school transcripts which indicate the last school's name and location along with grades and year of graduation. Size of graduating class, class standing, and test results are not normally on the transcript. Information on other high schools attended may be on the transcript but elementary schools generally are not. Similarly, colleges require transcripts from other colleges if transfer students are requesting course credit.

The selected junior college districts required American College Testing (ACT) scores for all entering students while the universities required Student Aptitude Test (SAT) scores. Some junior colleges however, use tests other than SAT or ACT. State colleges have a local option on whether to use ACT or SAT. Test scores are not required for third or fourth year transfers.

The name of the school currently attended, its location and the student's grade level (or year) are automatically part of any record. Grade point average may be developed by the institution but there are variations in the methods used to determine this average. Class standings are not determined. Major field or curriculum is maintained in each computer system.

On an exception basis, certain offices on the campus collect data from students for specific purposes. As an example, those students applying for financial aid must furnish financial resources data. Married students applying for on-campus housing may be required to furnish data regarding their immediate family. Those students required to pay tuition may have to furnish information not normally collected from other students.

Faculty counselors collect data from students during interviews. Most of this data is recorded on interview forms and filed in student folders separate from Administrative student files. Information on a student's educational and vocational objectives may be collected by counselors but is not required otherwise.

Each college or university campus at all levels is required to collect, on an annual basis, statistical ethnic origin and financial aids survey data for the U.S. Office of Education. The purpose of this data is to fulfill the requirements of Title VI of the Civil Rights Act of 1964. The state colleges, University of California, and University of the Pacific collect this data directly from the student on a voluntary basis during registration. A five to six percent non-response is generally experienced. This information is often on non-identifiable forms or cards, and in no instance appears on official student records. Junior colleges obtain this data only on students receiving college administered financial aids. Other attempts at collecting ethnic information have been made via visual sampling techniques or identification of surnames.

While each of the public institutions visited had machine usable student information files of some description, the amount of data stored in this form is dependent on the equipment available. The University of California has two administrative processing centers (Berkeley and Los Angeles), equipped with third generation computers. Computing facilities at the state and junior colleges covered the entire range from only EAM equipment to small or medium scale third generation computers. One campus used a terminal connected to the large computing facilities of a nearby private company. As might be expected from the above, data may be stored in punched cards, on magnetic tape or in disc files. While data stored in any of these forms may be considered machine-available, the problems of using this data in any one system are substantial. The large differences in formats, coding techniques, and storage methods would require extensive intervening processes prior to entering a single system.

During the interviews, it became increasingly apparent that considerable activity was taking place in the development of student information systems. The University of California indicated that the development of a university wide student information system had been initiated. The State Colleges plan a student information system although development work had not started at the time of the visit. On the other hand, the state colleges have developed a set of standard definitions and a preliminary list of data elements for the data processing and reporting operations of the colleges. This information is available in an Information Service Guide (4). The Los Angeles City Schools were developing a Student Information for a third generation computer with a planned implementation date of early 1969.

Further substantiation of this increasing activity was obtained through the Legislative Analyst's Office. There it was learned that both the State Colleges and University of California had consolidated their information systems groups, appointing new directors, and providing more rigorous delineation of responsibilities.

In addition to the above systems, Aerojet, through participation in the development of information systems for the California State Department of Education, is aware of a pupil-personnel information system available to all elementary and secondary schools on a voluntary basis. This system was placed in pilot operation in 1965 and presently is available as an approved package through regional data processing centers. The estimated elementary and secondary student coverage this year is about 700,000.

3. Review of TALENT and SCOPE Experience

In order to benefit from the experience of those who have been active in longitudinal studies similar to the proposed Student Flow Information System, the two most applicable projects, TALENT and SCOPE were investigated. Each represents a large scale attempt at following students over a long period of time.

Project TALENT was organized in the late fifties to develop an inventory of human resources, a set of standards for educational-psychological measurement, to indicate patterns of aptitude and ability which are predictive of success and satisfaction in various careers, and to provide a better understanding of the educational experiences which prepare students for their lifework. To accomplish these goals a probability sample of approximately 5% of the high schools in the country was drawn in 1960. The 400,000 students in grades 9 through 12 attending these schools were administered two days of educational-psychological tests and inventories. The results were stored in a computer data bank along with certain elements of information which would be used later to contact the student. The plan was to contact students at intervals of 1, 5, 10 and 20 years after graduation from high school. Thus, the followup studies were staggered so that each of the original grades was followed up in a separate year. At this time all of the grades have been contacted for the one and five year followup.

The followup for a given grade included three or four waves of a mailed questionaire, each wave spaced about one month apart. Returned questionnaires were processed by coding clerks, and a punched card prepared for each. These cards were used to control preparation of mailing labels for the next wave, including corrected addresses for those questionnaires returned by the post office.

Following the mailed survey, a 5 percent sample was randomly selected from the entire nonrespondent group. These individuals were sought out through field surveys conducted by regional coordinators and finally by the Retail Credit Company.

Response rates for the mailed followup one year after high school ran 69% for twelfth grade, 45% for eleventh grade, 43% for tenth grade, and 37% for ninth grade. The twelfth grade followup took place only one year after the initial testing while followup of each succeeding grade took place two, three and four years after testing. Hence, percentage returns decreased for each successive grade as the number of insufficient addresses increased and students began to lose interest in the program. During the five year followup, Project TALENT has maintained contact with approximately 35 percent of the original students.

Methods for contacting the nonrespondent sample differed slightly in each of the four one year followup studies. In the first three special surveys (twelfth through tenth grades), all questionnaires completed by regional coordinators or high school personnel were sent to the Retail Credit Company. In the fourth year the entire ninth grade nonrespondent study was assigned to the Retail Credit Company. Response rates in these special surveys were from 90 to 99 percent the first three years and only 73 percent the fourth year. The low rate of response in the ninth grade survey was due mainly to Retail Credit's inability to locate many of the nonrespondents within an imposed one-hour time limit for each questionnaire.

The overall cost per located nonrespondent including just the expenses of the schools, coordinators, and Retail Credit was approximately $7.00. However, if staff time were included for each of the first three surveys the costs would be appreciably higher. Frequently the cost of processing expense vouchers exceeded the expense itself. The cost per nonrespondent located in the study carried out by Retail Credit was $9.66.

One important area of TALENT experience was that of confidentiality. It was pointed out that difficulties were experienced largely because of a lack of understanding of the project. For this reason a good deal of effort was expended in publicizing the project and informing concerned persons of the philosophy of the project and protections which are exercised to restrict the use of the data. TALENT will not make available information which can be traced to individual persons. This is accomplished by eliminating all personal identity data when the data bank is used as a resource for certain research projects. An interesting aspect of the TALENT study was that 30% of the schools which were selected in California refused to participate in the study. This can partly be explained in that the schools were asked to test individual students. Many of the larger schools looked upon this task as an overwhelming burden.

SCOPE is a six year four state study of student decision making which seeks to find out how, when and why students make decisions about post-high school education and careers. SCOPE also seeks to determine the relative influence that parents, schools, and peers have upon the nature of those decisions.

SCOPE selected a sample of schools in California, Illinois, Massachusetts, and North Carolina and then in the spring of 1966 obtained data from the approximately 90,000 students in grades 9 and 12 by administering an academic ability test and a variety of questionnaires. Data received from the schools was then stored and attempts made to locate the students a year later.

The followup procedure used by SCOPE is considerably different than that used by TALENT. Rather than mailing directly to the individual, questionnaires were sent to the high schools or, in the case of the original twelfth graders, to their declared college choices. The schools were requested to check their rosters and obtain completed questionnaires for those located. For those not located by the schools, cards were sent to the student's home address requesting information on their location.

SCOPE's experience with the diversity of student flow and sample degradation was interesting. Approximately 300 high schools in four states were involved in the project at its inception. By the spring of the following year the ninth grade students had migrated to over 1500 high schools in all 50 states. The original ninth grade sample in California numbered approximately 9,000 students in 33 high schools. The following year these same students were attending 200 different high schools. Sample degradation compared closely with the TALENT experience. Of the original 46,000 ninth grade students SCOPE currently has contact with about 34,000.

SCOPE found it difficult, as TALENT also did, to gain the cooperation of certain large metropolitan school districts. SCOPE also cited difficulties with opposition from groups concerned with privacy. An important aspect of SCOPE was the feed-back of documents to participating schools which detailed the flow of students from that particular school. This was done as an inducement to gain the cooperation and interest of the schools and seemed to work well. This technique is included into the contractor's concept of the proposed Student Flow Information System for the same reasons.

In summary, both TALENT and SCOPE have goals related to but considerably different from those of the proposed Student Flow Information System. Nevertheless, their experience with regard to the techniques and problems associated with following individuals was extremely valuable.

Based upon the information summarized in the preceding paragraphs several basic decisions had to be made to best satisfy the purposes of the Student Flow Information System. These included what data elements should be collected, and how they should be collected. The following section shows how these questions were resolved and the reasons.

B. DECISIONS

1. Selection of Data Elements

The selection of data elements has been one of the more important and involved tasks of the study. As previously indicated the task began with creating a list of possible data elements and reviewing them with representatives of selected educational institutions. These on-campus interviews provided an estimate of the availability of information through the educational institutions.

The prime consideration in reviewing the information elements was the utility of each element to the proposed system and in particular how each element could provide additional insights which would be useful to legislative decision making. This emphasis on utility was naturally tempered by the ability to obtain information measured in terms of both feasibility and accuracy of the received information. In order to summarize this information a cross classification table was prepared which categorized the data elements according to their importance to the system and the effort required to obtain them. The result is shown in Table II-2. The necessity or desirability of data elements was based on their importance in identifying a student determining where he came from defining his socio-economic background, and assessing his academic ability and progress. Although specific considerations were involved in assessing the level of effort required to obtain each data element, the following definitions generally hold for the categories used in the table:

Generally Available Now - Available at nearly all institutions in some machine processable form.

Available at Modest Effort - Available at nearly all institutions but generally not in machine form.

TABLE II-2

NECESSITY VS EFFORT

Impossible	Last Elementary School	Spouse Income Size of High School Class High School Class Standing Student Income: Parental Support, Private Loans.	Parental Income	Post Educational Data
Available at Substantial Effort	Real Parents: Living, Marital Status. Parent/Guardian: Year & Place of Birth, Bilingual, Citizenship. Vocational Schools Completed Current Class Standing Number of Years in Last High School	Parent Education Brother & Sister Data Spouse: Education, Employed, Student. Children Birthdates Other High Schools Attended Educational Objectives Student Employment Education + Other Costs Student Income: Scholarships, Fellowships, G.I. Bill, etc.	Prior Name Parent/Guardian: Name, Relationship SAT/ACT Test Scores (Non-College Students) Grade Point Average (Other Colleges)	Ethnic Origin Parent Occupation
Available at Modest Effort	Current Address Selective Service Number		Country of Citizenship Marital Status SAT/ACT Test Scores (College Students Only) Other Colleges Attended: Name, Location, Degrees & Year Awarded.	Permanent Address Social Security Number Graduated High School & Year Full/Part Time Student Current Grade Point Average
Generally Available Now		State/Country of Birth	Date of Birth	Name Sex Last High School Attended ((State, County of Calif.) Current School Attending: Location, Grade Level, Major or Curriculum
	Marginal Value	Desirable	Extremely Desirable	Essential

Available at Substantial Effort - Generally not collected by by the institutions at this time.

Impossible - Experience indicates very high non-response rates and/or low reliability.

In general the educational institutions had little or no post educational data. Some alumni information might be available but aside from some small specialized studies, the location and activities of non-graduates was unknown. In addition to being beyond their current scope and financial capabilities, the conduct of large and continuing followup studies by individual institutions is impractical due to the extensive flow of students between schools.

As previously indicated ethnic origin is currently collected for the U.S. Office of Education but is not maintained in official student records permitting identification of individuals. Sensitivities in this area are considerable as is the problem of actually defining ethnic origin. However, this element is an extremely important key to defining socio-economic background and is considered essential to the proposed information system.

Likewise parental income is a prime descriptor of economic background. Investigations into studies attempting to obtain this information have found non-response rates of 35 percent and higher. In some instances this is due to a lack of knowledge by the student of their parent's income. In others the personal nature of the information and doubts as to how the data might be used cause failure to respond. It was also brought out that there appears to be a certain "halo" effect associated with questions of this nature creating serious doubts as to the reliability of the answers that are obtained.

The same problems relate to obtaining information on the student's own income or that of a spouse.

Since income appears virtually impossible to obtain, occupational information becomes increasingly important. Relationships can be established between at least broad levels of income or economic "status" and occupational categories. In addition, parental occupations are generally known by students and regarded as being less personal in nature. Thus, even though many institutions currently do not collect occupational data, it appears feasible for them to do so and imperative for the proposed information system if it is to fill the needs of the legislature.

While it is unnecessary to discuss each data element, reference to Table II-2 indicates that aside from only a few items all the information regarded as essential or extremely desirable is available through the educational institutions with modest effort. With a reasonably substantial effort nearly all elements of more than marginal value could be so collected. The main exceptions to this are income information and post educational data.

From the above analysis there was derived a final list of data elements (Appendix C) to be collected from students currently enrolled in an educational institution. Because those data elements (such as income) which were extremely difficult to obtain were equally difficult to obtain directly from the individual or through the institutions, the list was independent of the collection process. Post educational data was separated since it clearly had to be obtained directly from the individual through a followup process similar to that employed by TALENT. The selection of these data elements is discussed in a later section of this chapter.

2. How to Collect the Data

Having determined the desired information elements to be collected from students while enrolled in schools, the two methods that might be used to collect the data were examined. These two methods are either direct contact with the student or contact through the institution in which the student is enrolled.

Direct contact with the student will achieve success only to the extent of the willingness of the individual to reply. The response rate to be expected should be equivalent to the experience of SCOPE or Project TALENT. In making direct contact with the student, it must be recognized that certain information must also be obtained from the institution in which the student is (or was if the student has transferred) enrolled.

Collecting data from the student through the institution in which the student is enrolled offers the distinct advantage of using the established data collection processes. If the collection of the data is integrated into the regularly scheduled data collection operations such as registration, the credibility of the response as well as the percentage of response should in both cases be high. In addition, collection of data through the institution enhances the prospect of timely receipt of data at the processing center and in turn makes up to date information available to the legislature.

3. The Role of the Institutions in Data Collection

The utilization of data processing equipment for the receipt and storage of student information is already an accomplished fact at most educational institutions. The sophistication of this operation is directly related to the EDP capabilities of the school. There is a clear indication that the secondary schools, state colleges and the university system are developing or considering the development of information systems. At this time there is no indication that the junior colleges as a group, are considering such a development. It is felt that since the other three major segments of the educational population are moving toward centralized information, there is a strong possibility that the junior colleges will also. Even if this does not occur, discussions with representatives of the junior colleges indicated that, as a whole, they are quite advanced in the availability and use of data processing equipment. Review of a computer facility inventory compiled by the Board of Governors of the California Community Colleges showed that nearly all junior colleges have at least some EDP equipment. This is understandable when one considers that, in the community colleges, one of the major programs in continuing education is the data processing program. In any event, it is reasonable to assume that information elements from that segment of the educational population which is represented by the junior colleges will be available in machine processable form.

The students attending non-public schools must still be considered. Since the percentage of the educational population attending private schools at the secondary level is small (7%), and since it is felt that information can be obtained from some of the larger private secondary schools which possess data processing capabilities, it appears that a reasonable representation of private school students can be obtained.

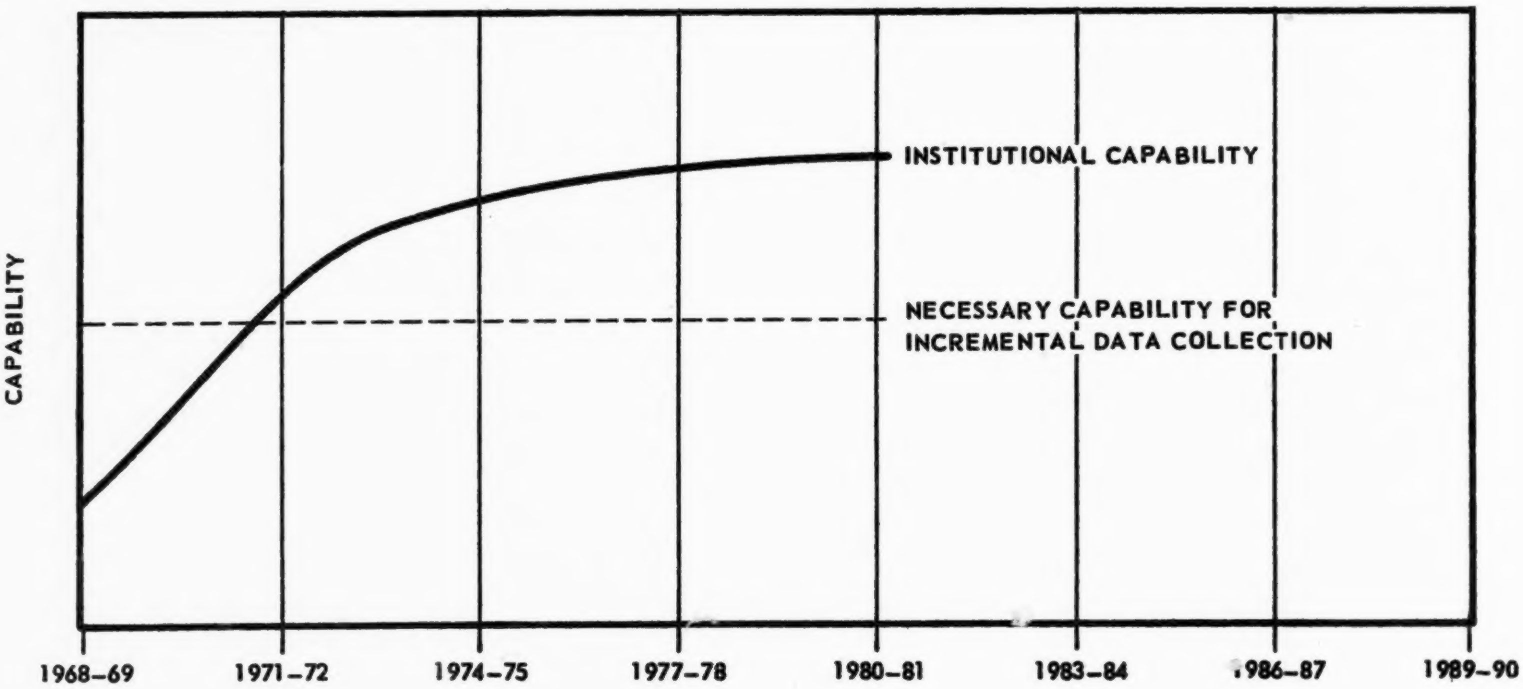
The private colleges and universities represent 12% of the population of the higher educational institutions. The significant proportion of students in this category attend private institutions which have rather sophisticated capabilities in data processing. When one considers that the University of Southern California and Stanford represent approximately 30,000 students and that an institution such as the University of the Pacific, with 3,757 students, is developing an extensive information system in the areas of admissions and registration, it is evident that a good representation of students from the larger private institutions can be obtained.

Considering the projected capabilities of the private institutions at the time when the proposed system is likely to be implemented, one is led to believe that a majority of the students in the private secondary and higher educational institutions can be included into the Student Flow Information System.

Information concerning the role private institutions play in California education is certainly an important asset of the proposed system and it is thus felt that every attempt should be made to include as many students from this group as is possible. It may be necessary and desirable to actively obtain a greater representation from the small private institutions in order to better determine their effect on the California educational system. This could be done by pursuing contracts where the institution would receive reimbursement for the incremental characters contributed to the Student Flow Information System. While this reimbursement probably would not be large enough to support the data processing activities necessary to provide such information, it may serve an an inducement to those institutions which are seriously contemplating such activity. If the reimbursement were offered while capabilities are developing in the private institutions it would serve to insure that all the information required by the proposed system was integrated into the developing information systems at each private institution.

The assumption is thus made that the institutional capability (measured in information element availability) for providing information will be sufficient to provide the necessary machine processable data at the inception of the Student Flow Information System. This implies that the system will be implemented after that time (1970 to 1972 in Figure II-1) when capability has exceeded the base capability necessary for providing data to a Student Flow Information System.

FIGURE II-1
EDUCATIONAL INSTITUTION
CAPABILITY VS TIME



If the assumed capability of the institutions is not present, two alternatives may be pursued. The first alternative would be to delay the implementation of the proposed system until the assumed capability is attained; the second would be to include the development of the proposed system into the systems development efforts of the educational institutions resulting in a joint effort. Clearly the latter alternative would be more economical since conversion programs and system modifications would be minimized.

4. Sample Size for Institutional Data Collection

The decision to utilize the information systems which are being planned and implemented within the educational system leads to the consideration of the sample size which will be used to collect data through the institutions.

Since the character of a Student Flow Information System implies the matching of information over time, the sampling problem becomes very difficult. If a 10% sample of the educational population were selected in one year, the problem of locating the members of the original sample in the following year is quite complicated. One alternative would be to try to locate the student through the educational institution as was done by SCOPE. There is a good chance that the student will have changed institutions and thus the required information would not be available, since in most cases, the institution will know the student left but will not know where he has gone. The SCOPE experience in this area indicates that the number of institutions which must be contacted will be quite large when compared to the original sample. The original SCOPE sample in California numbered approximately 9,000 students in 33 high schools. The following year these same students were attending 200 different high schools. It is evident that further pursuit of these students through the institutions may require that most of the institutions in the state participate in the study after a short time.

The flow of adults into and out of the educational system would be difficult for any sampling process to measure. Also, while the actual flow of students between institutions is unknown, it is recognized that it is both large and complex. As the amount of movement increases, sample degradation rates can be expected to increase requiring very large samples to maintain sufficient numbers by the time they reach young adulthood.

The nature of the proposed system and the information which is necessary from the system indicates that a rather large sample will be necessary. In order to arrive at sufficiently large cell sizes in particular areas of interest the original sample must be very large. When one considers the problems of retaining significant cell sizes and the problems of matching students over time it becomes evident that a very large sample must be used and that the students participating in such a sample will, in a very short time, be attending many different educational institutions.

A 100% system for students in school would provide complete detail of the student flow and clearly define populations of students who leave with specific educational backgrounds. Samples for followup of students graduating or dropping out of school could thus be allocated to insure a sufficient number from each population. In addition, the tracking of students within the educational system becomes a simple computer matching problem thus eliminating the problems of locating students.

The above reasons and the expense associated with sampling when compared with the costs of collecting data from the institutions indicate that a 100% sample of the educational population is the most practical and economical approach to the problem. In a later chapter of this report the cost of maintaining a 15% followup sample is developed. If one compares this cost with the cost of collecting a 100% sample, it is immediately observed that for the sample size which is required, the sampling of 100% of the students provides the optimal approach to the problem of collecting data.

Having decided to collect data through the educational institutions and having determined that a 100% sample for students in school is preferable, the data elements were divided into those to be collected at admission, registration, and graduation. This data was reviewed with approximately thirty representatives from additional high schools, junior colleges, state colleges and university campuses at a meeting called in San Francisco by the Chairman of the Joint Committee. Based on the discussion at this meeting, the availability of data through the institutions was verified and a final categorization determined. These lists of elements are shown in Appendix D.

5. Followup Considerations

The preceding sections have been concerned only with the students while they attend school. In Section A, the rationale and need for a followup process was presented. In this section some of the pertinent characteristics of that process (the data elements, the length of time a former student participates in a followup, the expected degradation rates, and the sampling techniques) are developed. The detailed description of the followup process follows in Chapter III.

a. Data Elements

Data collected during the followup will be somewhat different in nature from that collected through the educational institutions. Aside from certain identity, location and marital family data the information obtained will be directed at determining why the students left school, whether or not they are continuing their education elsewhere, and what they are doing. The latter includes information on occupational history, income, and aspirations.

Appendix E provides a list of followup data which is indicative of the types of data required. It can be expected that as the operation of the system progresses and as new hypotheses are formed, the information desired will vary relative to what has been learned. It is likely that some questions will differ according to the educational level at which the student left school. At times special studies may be desired for particular groups with certain characteristics. The determination of the specific information desired from various segments will be a function of the demands placed upon the system and will require continual analysis by the technical staff.

b. Length of Participation

In order for the followup data to be of maximum utility it should extend at least into young adulthood. An effective measure of what finally happens to an individual can only be obtained when a point of relative stability has been reached. Frequently the early years following education are dynamic years with gross changes in goals or direction. Certainly the current world situation and the prospect of military service adds to this initial instability. It can be expected that college graduates will generally reach a point indicative of their ultimate "success" more quickly than say a high school graduate or drop-out who doesn't go to college. The latter is usually younger and less mature and may not settle down to the business of "getting ahead" for several years.

For purposes of defining young adulthood, it seems reasonable to allow a college graduate five years after completion of his degree to reach a fairly stable point indicative of his "success". If we consider a "typical" graduate we could allow four years of high school, two years of military service, and five years of college. In other words, there is a total of 16 years beyond the entering high school freshman level to the "success" level. High school dropouts, graduates, and college dropouts should also be followed for 16 years beyond the freshman level. In this manner the different groups can be compared at the same point in life to determine the influence of the different paths they have followed. Certainly there is nothing sacred about the 16 year figure, perhaps it should be 14, 20, or even 30 years. However, 16 years would appear a minimum for those progressing through college. Until it could be shown that those following other paths reach a stable level more quickly it seems sound to follow them for the same period. Yet, the longer individuals are followed the greater the costs become and these increased costs must be compared with the value of data gained. Until more is known concerning the stabilization point and sample degradation rates it is not prudent to decide upon any other time period for followup studies.

c. Degradation Rates

A serious problem in all followup studies is the degradation of the sample sizes through time either through failure to locate individuals or failure on their part to respond. This problem becomes particularly acute over a long period of time when descriptive and comparative data are desired for each of several subpopulations. For this reason the original sample sizes are normally increased in order to provide the desired sample size and precision after the expected degradation.

Results of the TALENT study indicate that approximately a 70 percent response rate can be expected the first year of followup. Response rates for later years however, were not applicable since students were not contacted for 2, 3, or 4 years greatly increasing the number of erroneous addresses. Although there is little to go on it is felt that about a 10 percent degradation rate per year, after the first year's 30 percent rate, is a reasonable assumption. Table II-3 was calculated in accordance with these degradation rates and shows the remaining proportion of a sample expected after each year of followup.

Although TALENT provided no information along this line, degradation rates can be expected to vary between groups with different characteristics. Because of this and the fact that information will be desired on each group as defined by certain characteristics, a stratified random sampling procedure is required. That is, an independent random sample will be selected from each of the groups. Sample sizes would be proportional to the expected degradation rates.

TABLE II-3

EXPECTED PROPORTION, R, REMAINING IN A SAMPLE AFTER N YEARS OF FOLLOWUP

N	R	N	R	N	R	N	R
1	0.70	5	0.46	9	0.30	13	0.20
2	0.63	6	0.41	10	0.27	14	0.18
3	0.57	7	0.37	11	0.24	15	0.16
4	0.51	8	0.34	12	0.22	16	0.14

d. Sample Size

In order to determine the size of sample required for followup, the particular questions to be answered must be known as well as the precision desired in the corresponding estimates. An example of the type of question one might ask is: Of those male high school graduates who do not go to college and are in the lower socio-economic quartile and upper ability quartile, what proportion remain in the lower socio-economic quartile at young adulthood (12 years after high school graduation)? A similar question might be asked for any or all combinations of quartiles or for those who do complete varying amounts of college. The non-college-going high school graduates would form a control group for comparison with those with higher education and hence provide a good group from which to estimate sample sizes.

Approximately 76,000 California high school graduates in 1967 did not attend college the following year. Assuming half of these to be males there was an average of 2500 male students in each of the 16 cells of quartile combinations. The sample size required to be 95% confident that the estimated proportion be within plus or minus d (the precision) of the true proportion is then approximated by

$$n = m_o/(1 + m_o/2500)$$

where

$$n_o = 4P\ (1-P)/d^2$$

and P is the true proportion. This is largest when P is 0.5 and smallest when P is near zero or one. Using P at an intermediate value of 0.3, sample sizes were calculated for various values of d. This sample size was then extrapolated back to an initial sample size by dividing by 0.22 (from Table II-3). The results are shown in Table II-4.

TABLE II-4

ESTIMATED SAMPLE SIZES PER CELL

d Precision	n required Sample Size	m initial Sample Size	m as % of Total Cell Size	m/3 as % of Total Cell Size
0.03	681	3095	124	41
0.04	434	1973	79	26
0.05	297	1350	54	18
0.06	214	973	39	13
0.07	161	732	29	10
0.10	82	373	15	5

If reasonable interpretations are to be drawn and trends detected from estimated proportions within two-way classification tables, experience indicates those proportions should be precise to within at least ± 0.05. Table II-4 indicates that such precision for a single year's sample, after 12 years of degradation would require an initial sampling rate of over 50%. Most results however need not be based on data from exits of a particular year, but could utilize the combined data of three or four consecutive years. Combining data from three years, a precision of ± 0.05 should require a sampling rate of about 18%.

The average college graduate should reach young adulthood about 5 years after graduation. If the same questions as above were asked concerning college graduates, only a 5 year followup period would be involved. Table II-3 indicates about 46 percent of this sample should be intact. Thus, the initial sampling fraction need be only about half as big as for high school graduates or about 9 percent.

Since the average followup time over all groups should run somewhere between 5 to 12 years an average sampling rate of 12 to 13 percent is indicated. Recognizing the degree of approximation involved in unknowns such as degradation rates and refinement of cell or sub-population definition an average sampling rate of 15 percent of the exits from the educational population appears appropriate. Fifteen percent is meant as an indication of the number of individuals participating in the followup system. It is not meant to represent a constant percentage which can be applied to each stratum to determine the number of individuals to be selected. On the contrary, each stratum will most likely require a different percentage; the 15 percent being only an indication of the relative magnitude of the followup tasks.

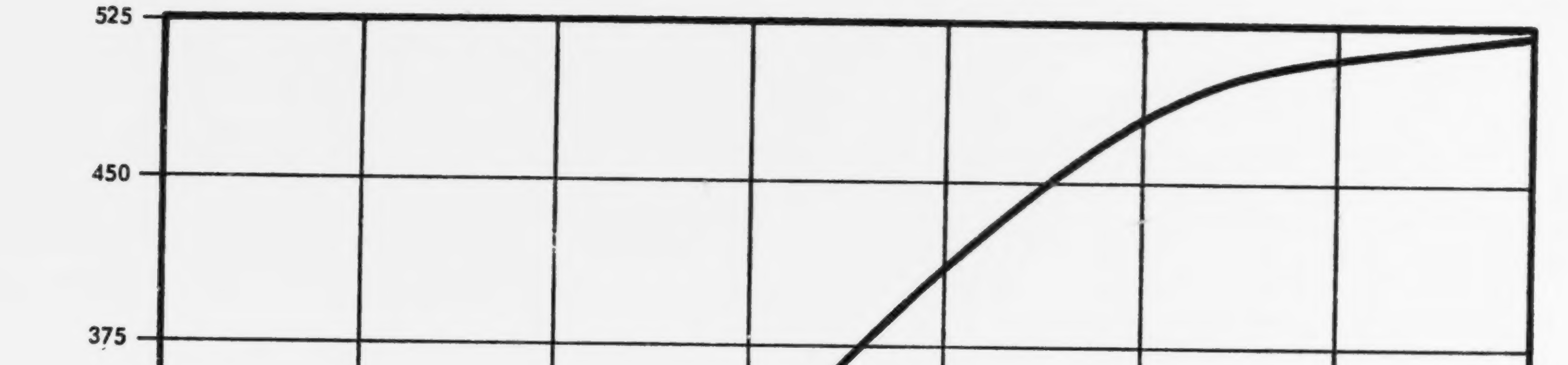
Table II-5 shows the number of individuals sampled each year based on this rate and the projected number of exits from the educational system. It also shows the cumulative number each year in the followup system using the above degradation rates. This is seen to stabilize at approximately 500,000 after 15 years.

TABLE II-5

FOLLOW-UP POPULATION

Year	Exits	15% Sample	System Size
1971	451*	67.6	67.6
1972	460	69.0	116.3
1973	475	71.2	162.1
1974	486	72.9	204.5
1977	538	80.7	320.3
1980	530	79.5	411.1
1983	560	84.0	482.3
1986	560	84.0	496.5
1989	560	84.0	503.3

* Exits between Fall 1971 and Fall 1972



III. SYSTEMS DESIGN

A. DEFINITION

In the previous chapter, the outline of the information to be collected from the student and the institution has been derived, likewise the populations from whom the information is to be collected have been established. These two factors make it evident that the collection, storage, maintenance and analysis of this information require a systems approach. Each task or effort associated with this proposed project must be carefully delineated and related to the preceding and succeeding task or effort. These tasks or efforts, taken together, have been called a Student Flow Information System.

Prior to discussing the detail of the proposed Student Flow Information System, two considerations need to be discussed: the staff which must control and direct the day-to-day design and operation of the system and the facility which performs the operation.

A project staff will be required to coordinate and direct the design, implementation and operation of the system. These individuals will also be responsible for budgeting and developing special studies to investigate various trends that become evident. This staff will consist of a project director, a limited number of technical assistants, and the necessary clerical help. The organizational implications of such a staff and its reporting responsibilities are presented in Chapter V.

A central processing facility will be required to meet the computer processing demands of the proposed system. It will consist of a medium scale computer and the appropriate personnel required to support the equipment. The responsibility for the computer facility will rest with one of the members of the project director's staff. Whether a separate computer facility should be established or the requirements should be met by a contract with a private service bureau, as well as the organizational implications of such decisions, is discussed later in Chapters IV and V.

Recognizing the possibility of oversimplification, the system can be described by the following sequential tasks:

> Collection of information from the student concurrent with the admissions and registration process: Data will be collected utilizing forms normally furnished by the institution, but modified or supplemented to include that information needed for the system and not normally collected in the admissions and registration process.
>
> Incorporation of all information needed by the Student Flow Information System into local or regional student information systems: Such systems exist in one form or another in the high schools, the colleges, and the universities. The State Department of Education has a student information system developed for high schools to use on a voluntary basis. Each junior college has a student-personnel system in some form. The state colleges and the universities are developing new student information systems.
>
> A maintenance processing cycle at the central processing center to take place after each registration: During this cycle, information received from the local or regional student information systems will be edited for errors and omissions and then added to the file of existing student records. The addition may be in the form of new student records or the updating of a record already in the file with current data.

Followup process: During the maintenance processing cycle, those students who were registered in previous semesters and who are not reflected in the current registration are identified. These students then become eligible to participate in the followup. A sample is drawn from the eligible students and these individuals become part of a continuing process which follows the individual into young adulthood. If there exist patterns of flow which warrant further investigation, specialized studies can also be conducted using the followup process and the data base to determine causes for such patterns.

The final phase of the processing cycle is the reporting operation: During this cycle, the system using preconceived programs will develop statistical and analytical reports for those students still in the educational system and for those that have left the system. These reports will be expected to furnish data needed for legislative decisions.

The above descriptions are presented as flow diagrams in Figure III-1 (Data Collection), Figure III-2 (Update Process), and Figure III-3 (Report Generation Process).

B. DATA COLLECTION

The data collection process is the more complex segment of the proposed Student Flow Information System. In it there will be two distinct types of data collection operations. The first is the collection of data from students in the educational institutions (either public or private). The other is the collection of data from individuals in the followup population.

FIGURE III-1

DATA COLLECTION

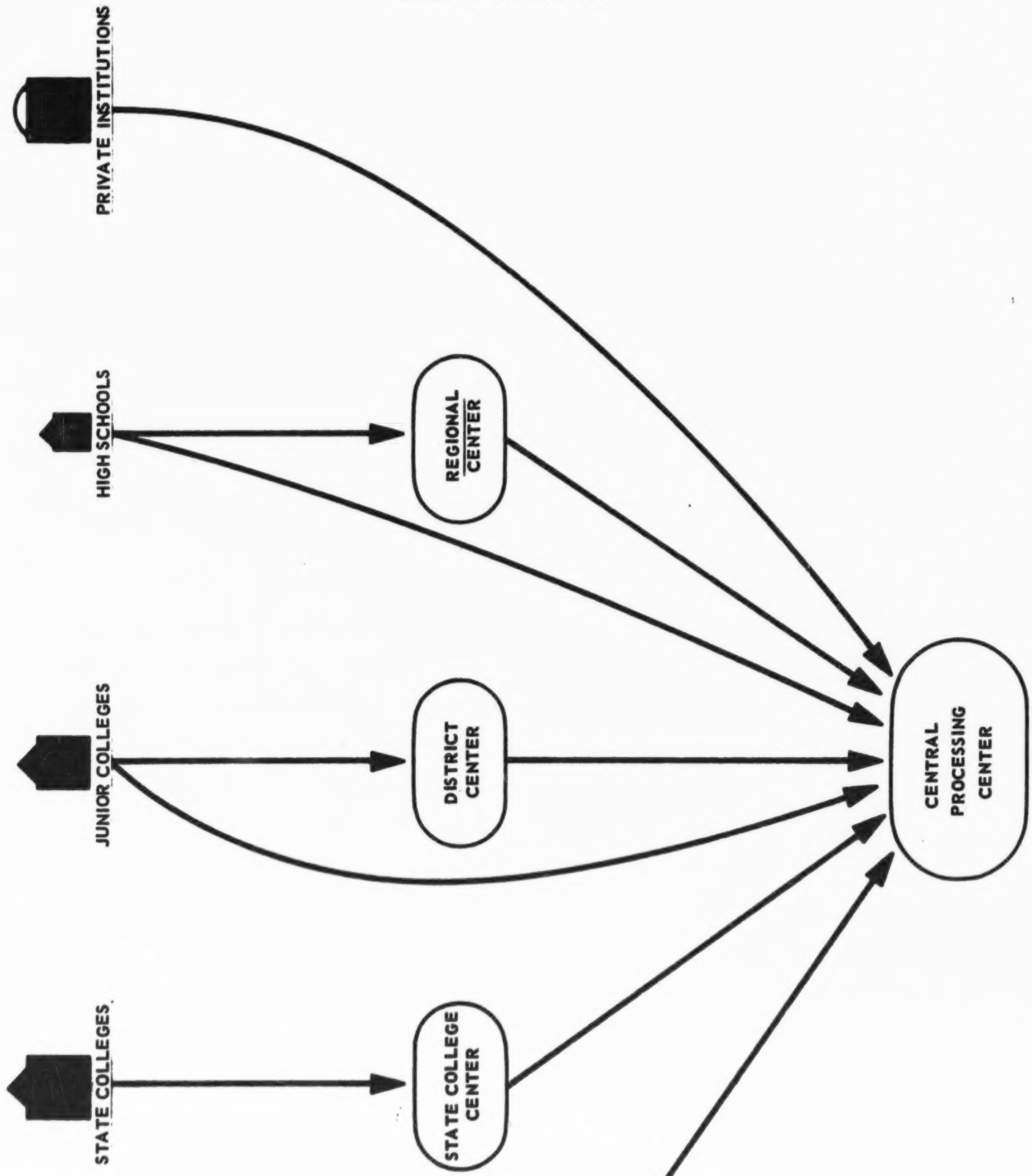

PRIVATE INSTITUTIONS
HIGH SCHOOLS
JUNIOR COLLEGES
STATE COLLEGES
UNIVERSITIES
REGIONAL CENTER
DISTRICT CENTER
STATE COLLEGE CENTER
UNIVERSITY CENTER
CENTRAL PROCESSING CENTER

FIGURE III-2

UPDATE PROCESS

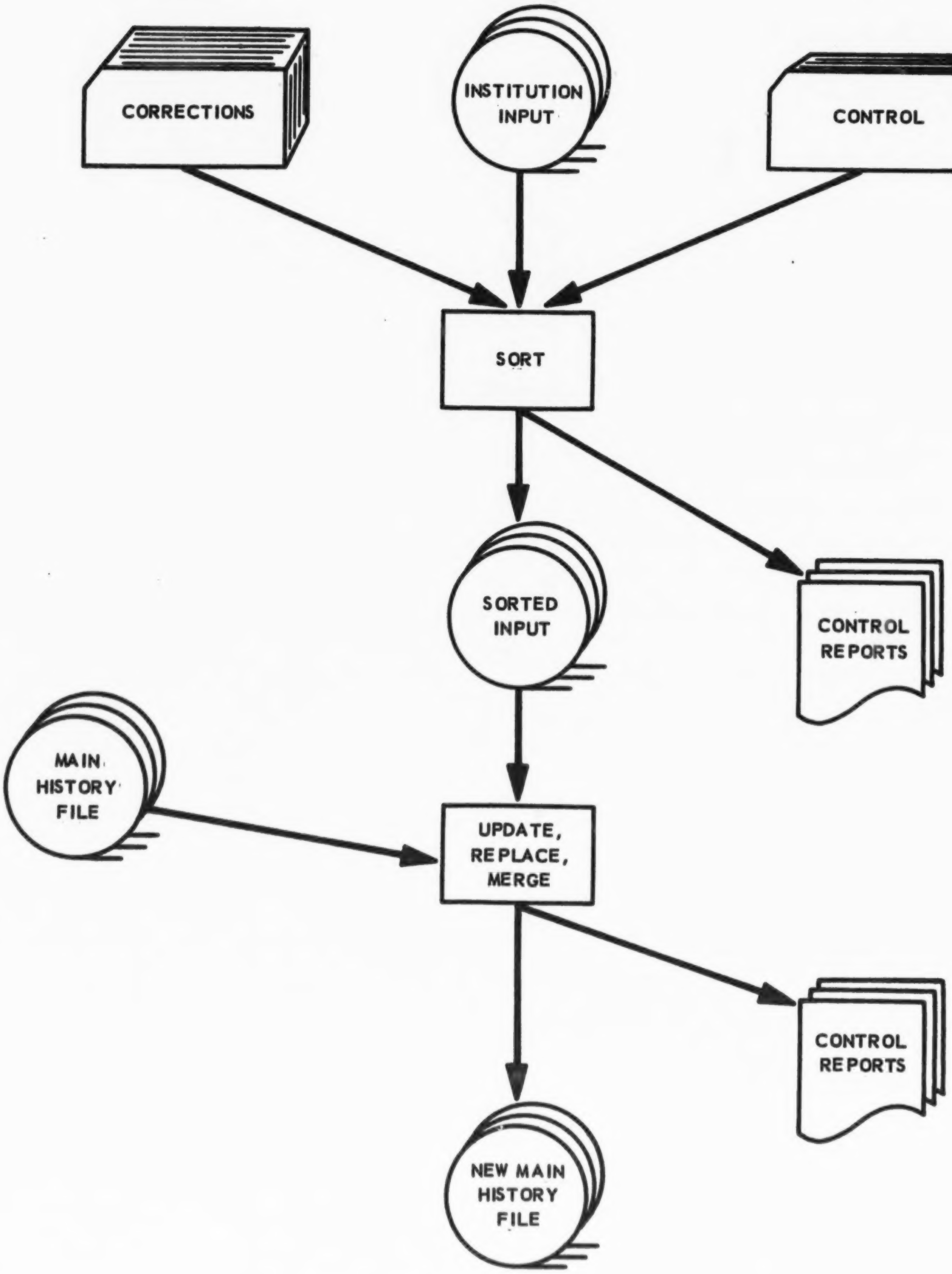
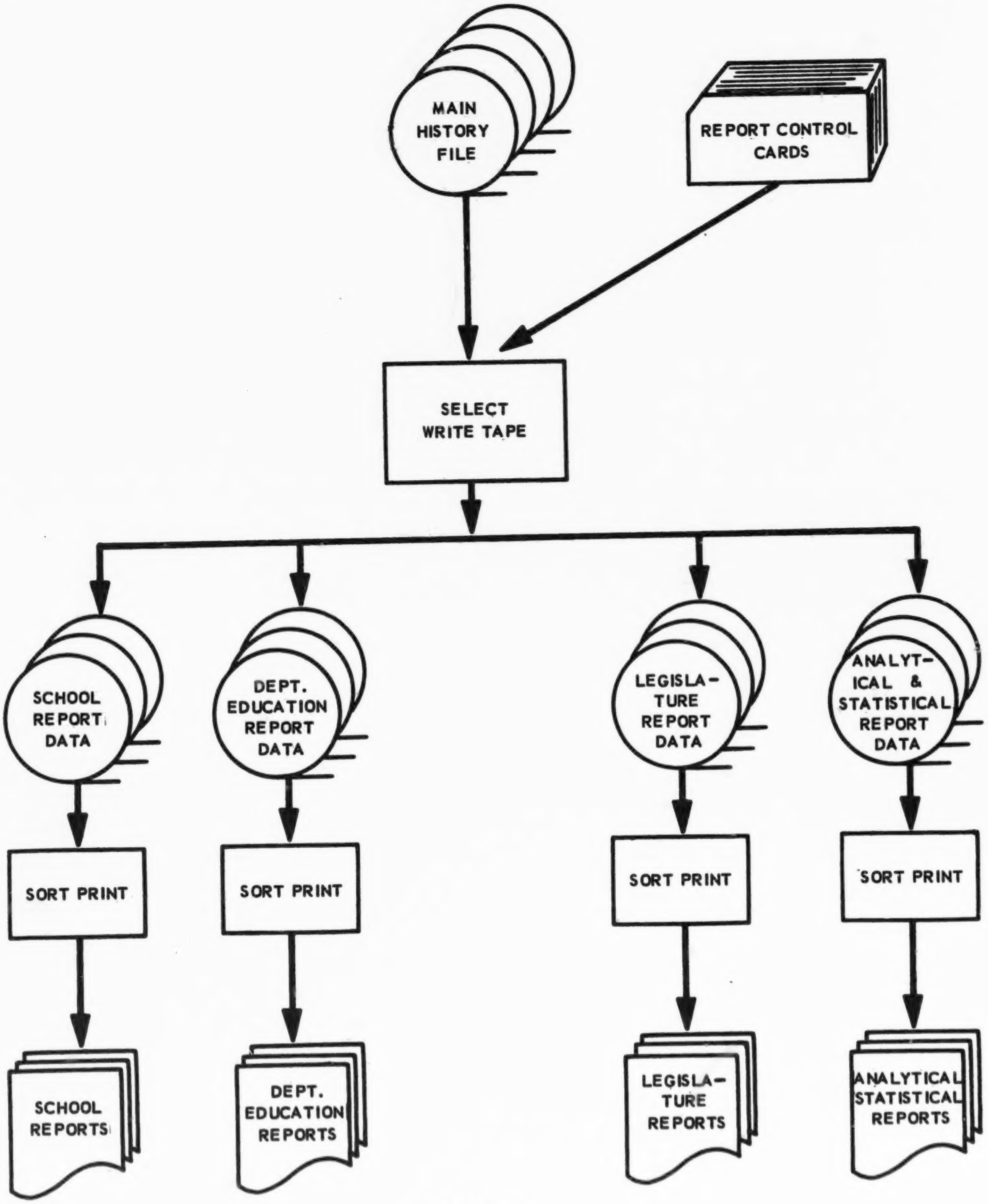

CORRECTIONS
INSTITUTION INPUT
CONTROL
SORT
SORTED INPUT
CONTROL REPORTS
MAIN HISTORY FILE
UPDATE, REPLACE, MERGE
CONTROL REPORTS
NEW MAIN HISTORY FILE

FIGURE III-3

REPORT GENERATION PROCESS


MAIN HISTORY FILE
REPORT CONTROL CARDS
SELECT WRITE TAPE
SCHOOL REPORT DATA
DEPT. EDUCATION REPORT DATA
LEGISLA-TURE REPORT DATA
ANALYT-ICAL & STATISTICAL REPORT DATA
SORT PRINT
SORT PRINT
SORT PRINT
SORT PRINT
SCHOOL REPORTS
DEPT. EDUCATION REPORTS
LEGISLA-TURE REPORTS
ANALYTICAL STATISTICAL REPORTS

1. Collecting Data from Students

The student will provide information to the proposed system by fulfilling the normal data requirements of his particular institution's admissions, registration and graduation processes. The instutions, during these operations, already collect certain of the data needed for the system and additional data can be included by modifying forms presently used for admissions and registration.

This information is processed by the educational institution and converted into machine processable data. These data will be available at regional centers in the case of the universities, state colleges, and for those secondary schools who participate in the Department of Education system. Data will be obtained directly from the private institutions and those secondary schools who have elected not to participate in the Department of Education system. The junior colleges have not, as yet, established a plan for a comprehensive information system. However, for reasons presented in Chapter II, it is assumed that machine processable data will be available.

It is envisioned that a limited number of data formats will be used. These formats will undoubtedly change to reflect different data recording mediums such as magnetic tape and punched cards. There may be separate formats for those regional centers who represent large portions of the educational population. The programs necessary to convert data to the formats required by the proposed system will be maintained by the regional centers or the institutions, not by the central processing center.

The information gathered by each institution for each registration will ultimately be forwarded to the processing center for the proposed system where it will be included into the data base.

2. Collecting Data from Former Students

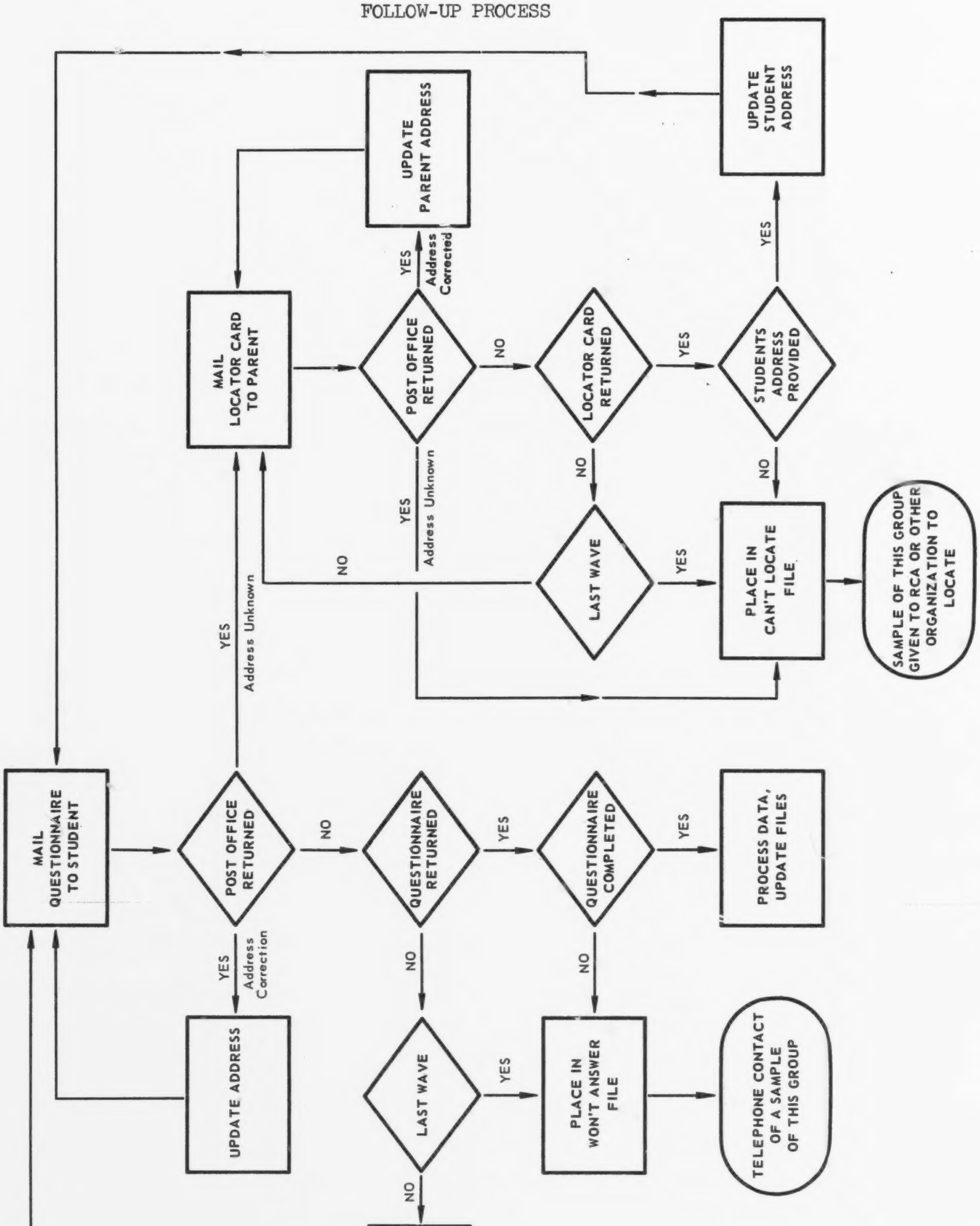
The difficulties to be encountered in the followup data collection operations have already been discussed. Every effort must be made to keep a current address for those in the followup populations. For instance, some form of newsletter, perhaps depicting the current activities or summarizing interesting results of the studies would be mailed annually to each participant. These would be sent with an address correction requested from the postal service. Returned addresses would be updated in the address file and the newsletter remailed to verify the new address. Questionnaires would be sent on an annual basis. In order to maintain as high a response rate as possible, questionnaires would be mailed in three or four waves, each wave being mailed only to the nonrespondents from the preceding wave. Attempts to locate individuals whose address is not up-to-date would be made by mailing cards to parents or other persons designated as usually knowing their location. Figure III-4 provides a schematic of the general followup procedures.

Questionnaires would be designed to permit direct keypunching (or perhaps optical scanning) as much as possible. Basic identity and address information will be preprinted on the questionnaire with a request for indicating any changes or corrections. This procedure reduces the effort required of the respondent, helps to ensure proper identity, and reduces errors by eliminating transposition of digits in social security numbers, etc.

An extremely important part of such a followup study is the data obtained from a sample of the nonrespondents. Although the effort or cost expended in locating and obtaining data from this sample of nonrespondents is usually very disproportionate to that for obtaining the original responses, it is in no way disproportionate to its value, at least until it has been verified that biases due to nonresponse are not serious.

FIGURE III-4

FOLLOW-UP PROCESS


MAIL QUESTIONNAIRE TO STUDENT
POST OFFICE RETURNED
YES
Address Correction
UPDATE ADDRESS
NO
QUESTIONNAIRE RETURNED
NO
LAST WAVE
NO
WAIT FOR NEXT WAVE
YES
PLACE IN WON'T ANSWER FILE
TELEPHONE CONTACT OF A SAMPLE OF THIS GROUP
YES
QUESTIONNAIRE COMPLETED
NO
YES
PROCESS DATA, UPDATE FILES
YES
Address Unknown
MAIL LOCATOR CARD TO PARENT
POST OFFICE RETURNED
YES
Address Corrected
UPDATE PARENT ADDRESS
YES
Address Unknown
NO
LOCATOR CARD RETURNED
NO
LAST WAVE
NO
YES
PLACE IN CAN'T LOCATE FILE
SAMPLE OF THIS GROUP GIVEN TO RCA OR OTHER ORGANIZATION TO LOCATE
YES
STUDENTS ADDRESS PROVIDED
NO
YES
UPDATE STUDENT ADDRESS

Appropriate methods for weighting results from the two response groups and obtaining unbiased estimates are given by Hansen and Hurwitz (7).

Project TALENT used about a five percent sample of the nonrespondents and found the Retail Credit Association quite effective in locating and obtaining responses from those selected. The same procedure is recommended here; however, when TALENT imposed a one hour time limit on finding an individual, the success rate dropped to about 73%. Because nearly a 100% response is required from those nonrespondents sampled in order to properly adjust for biases, this limit must be increased. Until further experience is gained, the optimum limit is unknown, but about two hours is suggested. In addition, biases must be determined for each of the subpopulations and because this reduces the individual sample sizes to a great extent, a 10 percent sample of nonrespondents is considered more appropriate.

C. CENTRAL PROCESSING

A central processing facility for the Student Flow Information System is the focal point for information entering the system and reports emanating from the system. The processing operation in itself is rather straightforward. However, the large amount of data to be manipulated during this processing requires complex and thorough built-in controls and edits to ensure the following:

- Non-duplication of processing
- Account for receipt of all input from the institutions
- Correct sequencing of processes
- Development of reports from the appropriate data
- Security of identity information for any individual

As indicated at the beginning of this chapter, the initial operation after receipt of all data from the institutions is the update cycle. The information received on new students is added to the history file as new records. The information received on students that have been in the system at other registration times will be added to information already in the file for this student. For the latter, positive identification must be made through the social security number and other means.

During the update cycle, identification will be made of those students that did not re-register. Such identification will indicate whether this is the first failure to re-register or just how many successive times the student has failed to re-register. Depending on the desired rule, a certain number of failures to re-register will place the student in an inactive student category for possible inclusion in the followup samples.

A second update is included in the followup processing and utilizes the data received in the replies to the followup questionnaires. Once a decision is made to followup a sample of a certain group of ex-students (the group is identified through some one or several commonalities), individuals are selected from the group based on a specified sampling technique. The system provides address labels and flags those that are in the sample. Data from the replies are used to update the existing records and provide data on nonrespondents. A sample of the latter is selected for intensified followup and the results of this are added to the files.

With the history file updated by both the new registration data and the followup information, the processing center begins the reporting cycle. During this cycle, four general classes of scheduled reports are prepared from production programs prepared in advance. These include reports as follows:

Legislative Reports - The greatest number of reports will be furnished to the Legislature. Each will contain vital statistical information important to legislative decision making in such areas as:

- Student financial aid
- Institution financial support
- Student migration
- Student performance statistics
- Utilization of higher institutions
- Length of time in school

These reports will provide summary statistics on graduation rates, dropout rates, transfers between similar as well as different types of institutions, etc. Tables will be provided comparing these statistics for students from different socio-economic backgrounds, from different regions of the state, for single and married students, etc. Data from followup studies will be used to provide comparisons of income, occupation, vocational satisfaction, etc., for individuals completing one, two, three, or more years of college. The amount and variety of useful information for legislative reports seems almost unlimited.

In addition to the regularly scheduled reports discussed above, the system will be capable of responding to special requests by the Legislature or other approved educational groups. It is anticipated that, as accumulation of data progresses, the handling of special requests will become a substantial part of the reporting task.

While the above are examples of possible reports which can be obtained from the system, it must be remembered that the value of the reports is dependent on the exactness with which the data elements are defined, the validity of the data received from the student, and the promptness of the institutions in transmitting the data to the processing center. Without question, those receiving some of the initial data reports from the system will wonder why additional information was not gathered from the student. In reply to this, it must be emphasized that a minimum set of data was selected as a more palatable beginning. Additional data elements may be added and probably will be added later at the request of the institutions.

D. IMPLEMENTATION OF SYSTEM

Before definitive methods of implementing the Student Flow Information System can be presented, there should be further definition and development of the system. Further, there should be some indication or decision that the college or university student information systems under development will be aligned with the proposed Student Flow Information System.

As an estimate, it appears that it will take the better part of 18 months to have the system designed, programmed and ready for implementation. The volume of data to be transmitted and processed when the system is fully operational is such that full implementation should not be attempted initially. Initial implementation should be made on a limited basis, using selected high schools and junior colleges in the same geographic area. Then as this operation smooths out, additional institutions can be added. During such an implementation, the data reports will not be complete, but will provide a basis for analysis and modification. Proceeding in this way will add 12 months to the time required to bring the system into total operation.

Closely related to the discussion of implementation is the alternative of phasing the implementation of the Student Flow Information System. It is evident that there will be a certain amount of phasing required due to the inability of certain segments of the educational population to furnish the necessary information. In particular, the junior colleges and the private institutions might fall in this category.

The main concern which is related to the phasing of implementation is that of utility of the data during the time period required for the phasing. For example, if it was necessary to phase the high schools or junior colleges into the proposed system over a three-year period, the utility of the proposed system during that period would be questionable. In particular, the system would not be able to provide pertinent information concerning the flow of students with any degree of accuracy. In the final analysis, it would appear that it would be better to wait three years to implement the system since the expense required to operate the system would not be reasonable, when compared with the utility of the system.

IV. COSTS

The costs of the proposed Student Flow Information System are developed according to the resources required for design, operation and staffing of the effort. The discussions of cost are then summarized to provide a total cost for each fiscal year. Included in the summary is a discussion of the accuracy of the costs and justification for the minimum and maximum estimated costs presented.

A. DESIGN AND PROGRAMMING

It is envisioned that the most difficult portion of the design and programming of the proposed system will be the coordination of the various groups within the educational system. To obtain the data required by the system, it will be necessary to spend many seemingly unproductive hours of conferences, meetings, travel and negotiations. These hours will undoubtedly provide very important results in the accuracy and reliability of the data which is ultimately received. One of the most useful results of these coordination efforts will be the establishment of standard codes which will enable meaningful interchanges of information between institutions and, even more important, the ability to obtain and combine information from each educational group.

For the reasons mentioned above the system design cost is estimated to be $150,000 over an 18-month period, beginning in Fiscal 1969-70. This estimate is based upon the contractor's experience with similar large scale projects and an estimate of the difficulty of the coordination task.

The cost includes the expenses which will be incurred during the negotiation and coordination phases of the project as well as the formal detailed analysis of the proposed system.

The cost of programming the system is estimated to be $150,000 over a 18-month period, beginning in Fiscal 1970-71. This was determined again from the contractor's experience in similar projects. Normally, programming costs are approximately double the costs of system design. However, the expected amount of coordination in the design phase is the basis for the one to one ratio. It is felt that the system does not represent a complex programming task but one which involves massive amounts of data. For this reason the emphasis, as reflected in the costs, has been placed upon the coordination of the coding structures and similar efforts aimed at choosing the most effective strategy for solving the problem. The programming considerations should remain straightforward regardless of the strategy chosen.

B. OPERATION

The cost of operating the proposed system depends heavily upon the amount of data to be collected and maintained. The amount of data in turn depends upon the size of the educational population. For these reasons, population data were developed and are presented in the form of cost factors. From these cost factors, data collection and central processing costs were developed.

1. Educational Populations

In order to provide approximate costs for the design, development, and scheduled processing of the proposed Student Flow Information System, certain basic data was required relating to the estimated number of students to be encompassed in the system and the amount of information (characters of data) to be retained about each. The estimated number of students in the system is directly related to the number of students enrolled in the California secondary and higher educational institutions, the annual increases to this enrollment, the number of students that leave (exit) the educational system prior to and at graduation, and the sampling techniques to be used in following students.

The publication - Total and Full-time Enrollment, California Institutions of Higher Education, Fall 1967 (8) was used to develop percentage factors relative to the distribution of the total Fall 1967 enrollment among the grade levels of public and private high schools and junior colleges as well as total enrollments within the public and private colleges and universities. The total enrollment in these institutions (exclusive of nonsectarian secondary schools) and the percentage factors were shown in Table I-1.

Yearly enrollment projections of students in public school grades 9 through 12 were obtained from California Population 1967 (2). These projections were multiplied by 1.07 (the ratio of the total number of students in public and private high schools to the number of students in public schools for the year 1967 - Table I-1), to obtain total secondary school enrollment projections through 1980 (Table IV-1).

A third publication, California's Needs for Additional Centers of Public Higher Education (1), provided a chart of projected full-time enrollments in California institutions of higher education through 1980. This projection was multiplied by a factor of 1.877 (the ratio of total to full-time enrollments in the Fall of 1967) to obtain a projection of the higher-educational total enrollments through 1980 (Table IV-1).

The secondary and higher educational populations from 1971 through 1989 show essentially constant populations in projections beyond 1980. This is based on the current Department of Finance projections of a reducing high school population starting about 1979. Whether this trend will continue

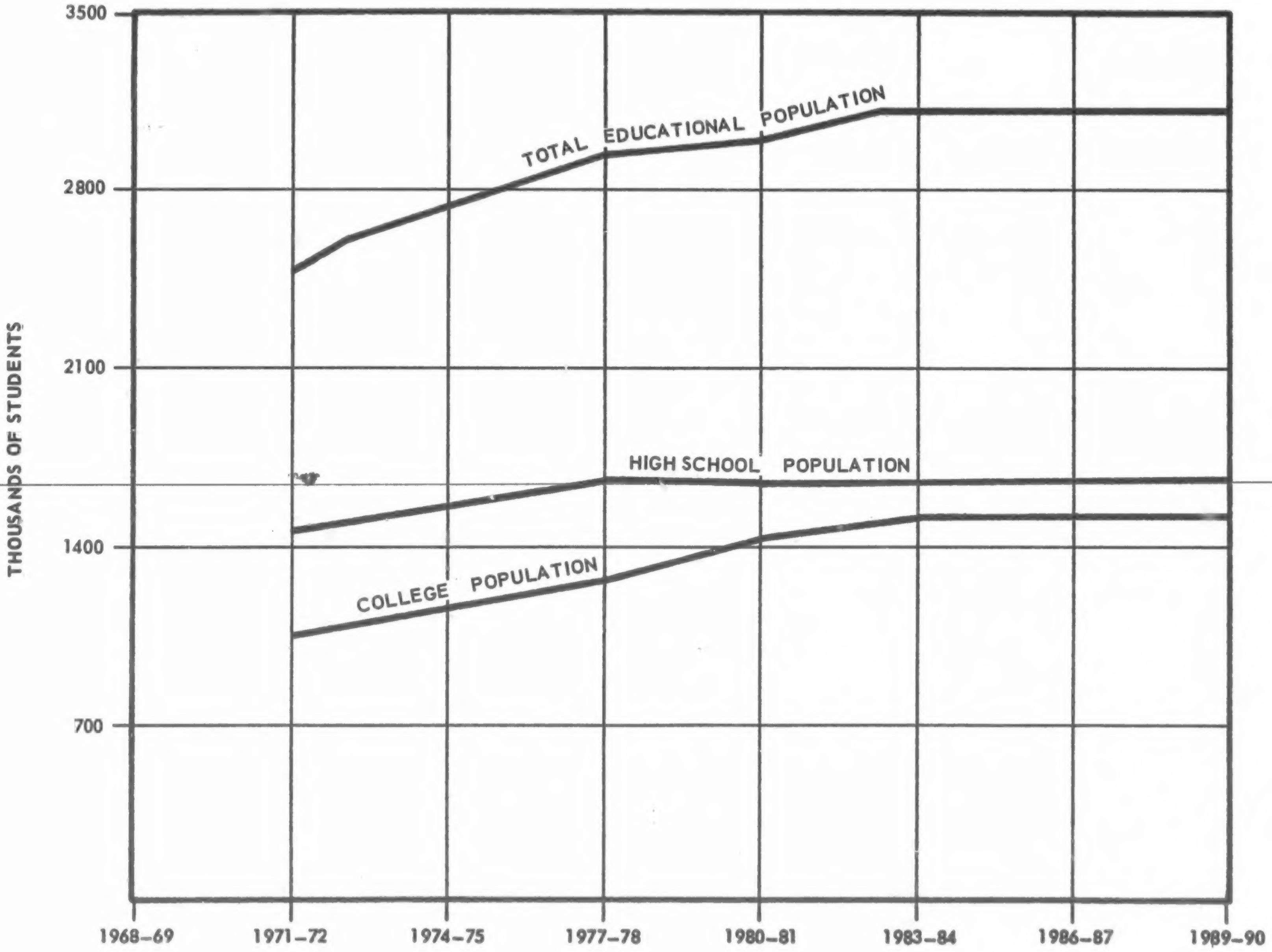
TABLE IV-1

PROJECTED POPULATION DATA
(THOUSANDS OF STUDENTS)

EDUCATIONAL POPULATION
(STUDENTS IN HIGH SCHOOLS, COLLEGES, AND UNIVERSITIES)

Year	High School	College	Total	Annual Increase	New Entrants High School	New Entrants College	New Entrants Total	Exits
1971	1469	1051	2520					
1972	1502	1092	2594	74	481	44	525	451*
1973	1537	1134	2671	77	492	45	537	460
1974	1568	1177	2745	74	502	47	549	475
1977	1646	1297	2943	59	527	52	579	520
1980	1597	1419	3016	14	511	57	568	554
1983	1600	1500	3100	0	512	60	572	572
1986	1600	1500	3100	0	512	60	572	572
1989	1600	1500	3100	0	512	60	572	572

* Exits between Fall 1971 and Fall 1972



or whether the population will again start to increase is a complex demographic problem beyond the scope of this study. However, it was felt that projecting a continuing downward trend could easily result in underestimating system costs. Hence, high school projections beyond 1980 were made on a stabilized basis. Similarly, since the high school population is a prime feeder of the higher educational institutions, the college population was also assumed to stabilize starting about 1981.

Each year, the educational population in the secondary schools and higher institutions is increased by the new high school freshman class, and the high school and college transfers coming from out-of-state. The references used above also provided data on out-of-state transfers into the junior colleges and state colleges, the universities and the private institutions for the 1963 school year. Since high schools and junior colleges are both community oriented, the annual entrants, exclusive of beginning freshmen, for high schools can be assumed to approximate that of the junior colleges which was 6% in 1963 (junior college entrants divided by the 1963 junior college full-time enrollment). To be on the conservative side and as a result of an investigation at one local high school, the transfer rate of 6% was adjusted downward to 4%. Thus the new entrants to high schools each year are estimated to be 32% of the projected high school enrollments (4% transfers from out of state plus the 28% from Table II-1 representing the percentage of freshmen in the total annual high school population).

New entrants at the college level include only transfers and first time freshmen from out-of-state since most of the entrants are from California High Schools and hence already part of the educational population. The reference data indicated approximately 4.7% of the 1963 total, full and part time higher educational enrollment to fall in this category. This rate was rounded down to four percent on the assumption there would be some decrease in the out-of-state enrollment.

Based on these percentages, Table IV-1 projects 525,000 new entrants to the total California educational population in 1972 increasing to 581,000 in 1978 and stabilizing at 572,000 after 1983. The number of students exiting the educational population each year, also shown in Table IV-1, were obtained by subtracting the annual enrollment increase from the number of new entrants.

While the educational population projections have been developed using the best data available and adjusting that data with factors that appear logical, these population projections must be recognized for what they are. For instance, the projections were calculated assuming that the distribution of students in grades 9 through 12 will remain as in 1967. Nonetheless, these projections appear to represent the trends, at least during the early years, and cost estimates based on these should be expected to be reasonably accurate.

2. Data Collection

The data collection costs are developed in two general areas - collecting data from institutions and from individuals. The former represents the tracking and matching of students actively engaged in pursuing an education, the latter represents the collection of questionnaire information through the follow-up system.

a. Data from Institutions

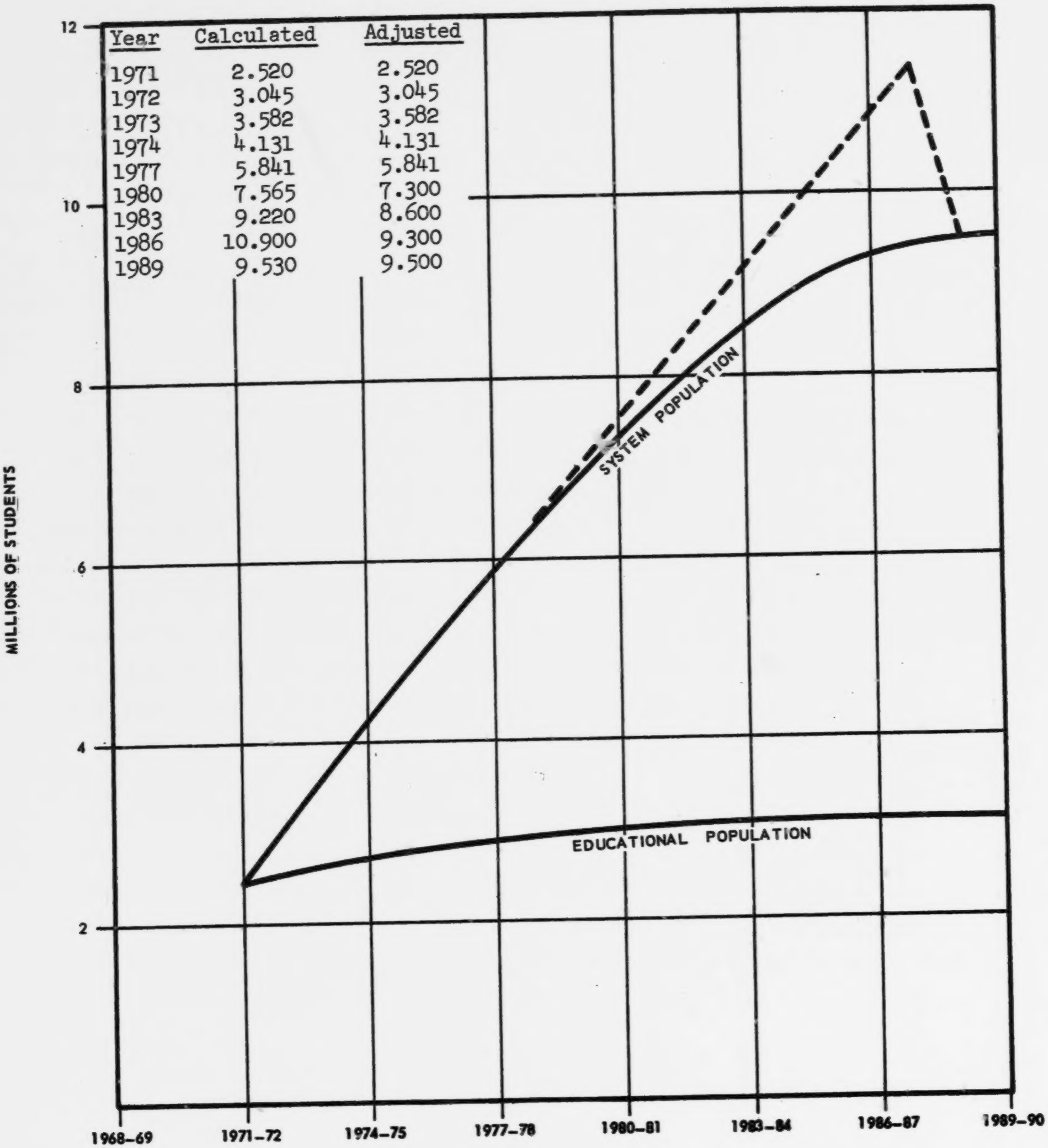
In order to determine the magnitude of the task of collecting data from the educational institutions, it was necessary to obtain an estimate of the total number of students who will be eligible to contribute information to the proposed system. When the proposed system is implemented (assumed to be fiscal 1971-72), the population in the system will equal the total educational population. These records will be maintained for several years while the student is attending school and following a student's leaving the educational process. The calculated data in Figure IV-1 is thus based on maintaining records for sixteen years and adding records for the new entrants each year. The system population is then shown to increase from 2.52 million in 1971 to a maximum of 11.46 million in 1987. At that time the system would drop the original 2.52 million students and pick up 572,000 new entrants (Table IV-1). After that the new entrants and number of records dropped per year became approximately equal, stabilizing the system population around 9.5 million records.

It is recognized that when the system is implemented many of the students will already be college seniors, graduate students, etc., and that follow-up would be something less than 16 years, perhaps only 3 or 4 years. Hence, the system population can be expected to approach the 9.5 million figure in a smoother fashion as indicated by the adjusted values and the solid curve of Figure IV-1. The adjusted value however was not allowed to drop off too rapidly in the early years since new entrants can be expected at the college level who attended California schools previously but dropped out for a period of time. While the system would have records on such students after being in operation for several years, it would not have these records during the initial years.

The volume of information to be collected from those students who contribute to the system is determined by the number of characters of data which are required at each admission, registration and graduation and the frequency of these events. It was found that the volume of information varied significantly between the high school and higher educational population but not within either group. Thus, the community colleges, state colleges,

FIGURE IV-1

SYSTEM POPULATION



Year	Calculated	Adjusted
1971	2.520	2.520
1972	3.045	3.045
1973	3.582	3.582
1974	4.131	4.131
1977	5.841	5.841
1980	7.565	7.300
1983	9.220	8.600
1986	10.900	9.300
1989	9.530	9.500

and universities were grouped into a higher educational population for the purpose of obtaining an estimate of the characters of information for any one year. The important factor thus became the frequency of the various activities within each of these populations. The following assumptions were based upon discussions with various educational institutions and pertinent statistics obtained from the references given in the section entitled Educational Populations.

(1) The entire population would submit admissions data the first year; 40% of the high school population and 35% of the higher educational population would submit admission data in subsequent years.

(2) Twenty percent of the total population would submit graduation data each year.

(3) The high school population would experience two registrations per year and the higher educational population three per year.

These assumptions and the characters of data to be collected at admission, registration and graduation permit the calculation of the total number of characters to be collected each year. The volume of information to be collected by year is presented in Table IV-2.

TABLE IV-2

TOTAL CHARACTERS IN MILLIONS

Year	Admission	Registration		Graduation	Total
		High School	College		
1971-72	300	510	565	30	1405
1972-73	118	523	587	31	1259
1973-74	121	536	608	32	1297
1974-75	124	550	630	33	1337
1977-78	134	575	700	35	1444
1980-81	136	595	760	36	1527
1983-84	140	595	810	37	1582
1986-87	140	595	810	37	1582
1989-90	140	595	810	37	1582

Since the collection of information from the institutions assumes that institutional student information systems will provide the mechanism through which the proposed system will obtain data, a measure of the increase in the data collection tasks in the institutions is needed so that the additional expenses which are imposed by the Student Flow Information System can be isolated. The additional characters of data which must be collected, presented as incremental characters, are shown in Appendix F. The total incremental characters for admission, registration and graduation are summarized below in Table IV-3.

TABLE IV-3

INCREMENTAL DATA COLLECTION CHARACTERS

	High School	Junior College	State College	University
Admission	13	40	40	40
Registration	19	84	84	84
Graduation	0	0	0	0

The information in Table IV-3, when multiplied by the appropriate populations, (Table IV-1), provides an estimate of the additional volume of information which must be added to the institutional student information systems if these systems are to supply the data required by the proposed Student Flow Information System. The additional volume of information is shown in Table IV-4.

TABLE IV-4

INCREMENTAL CHARACTERS IN MILLIONS

	Admissions		Registration		
Year	High School (13 Characters)	College (40 Characters)	High School (19 Characters)	College (84 Characters)	Total
1971-72	7.7	14.8	56.0	254.0	332.5
1972-73	7.9	15.4	57.0	264.0	244.3
1973-74	8.1	16.0	58.0	274.0	256.1
1974-75	8.2	16.6	59.0	284.0	367.8
1977-78	8.5	18.2	62.0	312.0	400.7
1980-81	8.3	20.0	61.0	341.0	430.3
1983-84	8.3	21.0	61.0	364.0	454.3
1986-87	8.3	21.0	61.0	364.0	454.3
1989-90	8.3	21.0	61.0	364.0	454.3

The cost of collecting data was estimated on the basis of the effort required to keypunch and verify the characters of data which are to be collected. A rate of $6.25/hour was used to estimate the labor, machine time, supervisory effort and general overhead associated with this process. Based upon the type of data to be used, it was assumed that 7000 characters of data could be keypunched per hour. When verification of the keypunched data was included, a combined rate of $1.78/1000 characters of information was obtained. By applying this rate to the total and incremental characters of information to be collected, (Tables IV-2 and IV-3), an estimate of the data collection costs of the Student Flow Information System was obtained. This estimate is presented as Table IV-5 and the cost of both alternatives is shown graphically in Figure IV-2.

TABLE IV-5

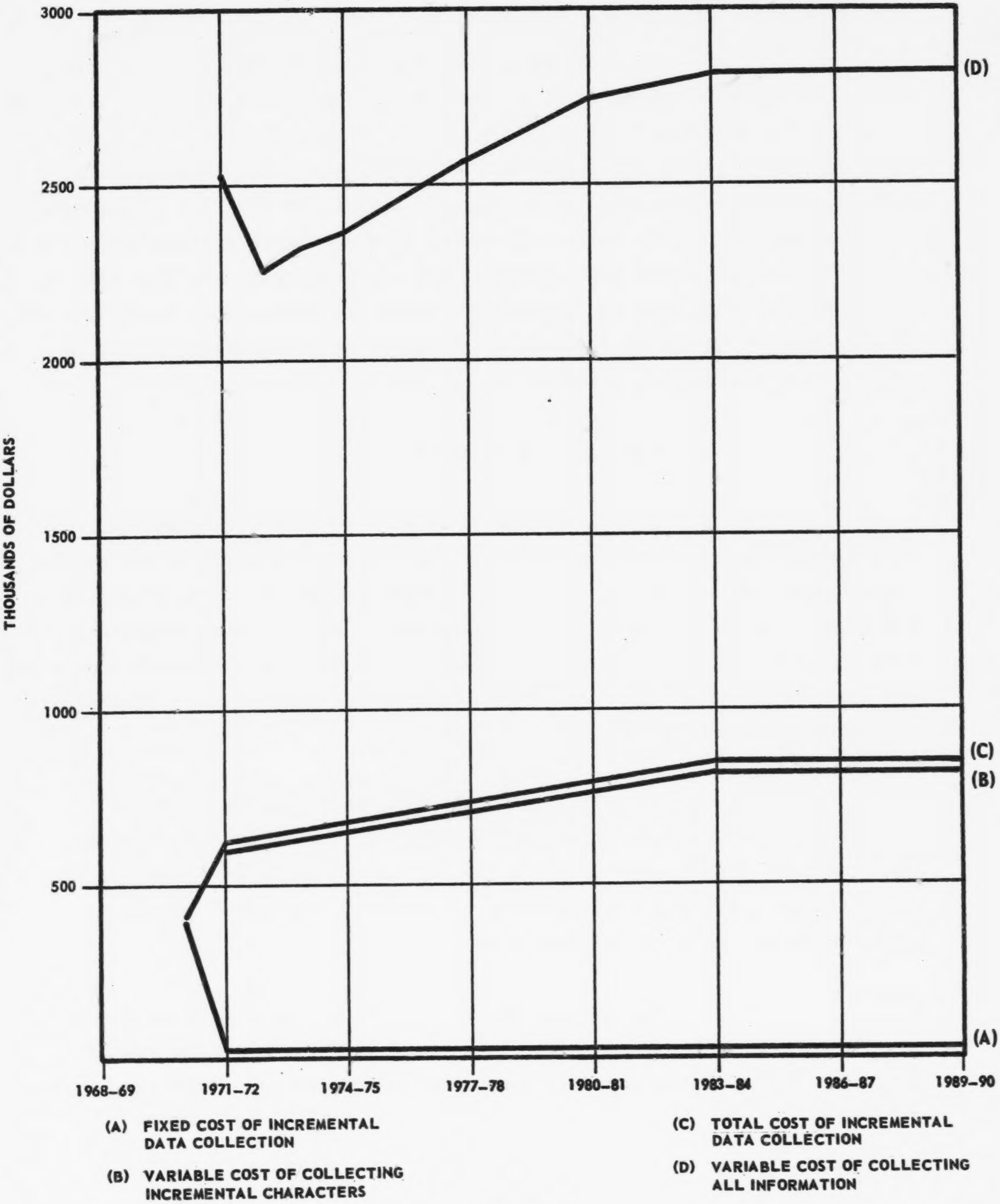
COLLECTION COSTS
(IN MILLIONS)

Year	Variable Cost of Incremental Data Collection	Fixed Cost of Incremental Data Collection	Total Cost of Incremental Data Collection	Cost of Total Data Collection
1970-71		0.400	0.400	
1971-72	0.591	0.020	0.611	2.51
1972-73	0.612	0.020	0.632	2.20
1973-74	0.633	0.020	0.653	2.27
1974-75	0.656	0.020	0.676	2.36
1977-78	0.714	0.020	0.734	2.57
1980-81	0.768	0.020	0.788	2.74
1983-84	0.810	0.020	0.830	2.82
1986-87	0.810	0.020	0.830	2.82
1989-90	0.810	0.020	0.830	2.82

The alternative of incremental data collection has associated with it certain fixed costs of integrating the proposed system into the various institutional information systems. These costs (such as re-designing forms, writing conversion programs, and processing information yearly), have been shown by year in Table IV-5 under the columns "Fixed Cost of Incremental Data Collection."

FIGURE IV-2

COLLECTION COST



In order to provide the proper perspective for comparing the alternatives of total vs incremental characters the total character collection cost (Table IV-5) was developed by calculating the cost of keypunching and verifying all characters of information. The total character collection cost is not meant as an accurate measure of the cost of pursuing that alternative. It is however developed as a yardstick against which the economies of incremental data collection can be compared. For this reason, no attempt has been made to increase the total collection cost to reflect the difficulties of obtaining accurate and complete data documents for use as input to the keypunching process.

b. Data from Individuals

The costs associated with obtaining information from individuals through the follow-up system are directly related to the sample size chosen. Once a sample is drawn the costs can be estimated according to the tasks required in the follow-up procedures - mailing questionnaires, processing returned documents, and locating non-respondents. Computer processing and staffing costs are included in the estimates for those items in their respective portions of this report.

By applying the 15% sampling rate, developed in Chapter II, to the projected number of exits from the educational system, the number of students who comprise the follow-up population (Table II-5) was obtained. The follow-up population represents the basis for the costs which are incurred in these areas:

Mailing and Materials - Questionnaires or newsletters were assumed to require one page printed on both sides involving printing costs of $155 per 5,000 copies. Envelopes were priced at $23.90 per thousand

and mail room processing at 1.8 cents per piece. These three items amounted to 7.3 cents per item. Postal rates run 6 cents regular or meter stamp, 8 cents business reply, and 10 cents per item returned with address corrections. The following assumptions were made: one newsletter would be mailed to each individual per year, one address correction and followup mailing would occur for 30% of the followup students and an average of two questionnaire mailing waves would be required. Using these figures, a mailing and materials rate of 47.4 cents per individual was obtained. An additional postage cost of 8 cents occurs for each questionnaire returned.

Processing of Returned Documents - Returned questionnaires were assumed to require minor coding at the rate of 30 questionnaires per hour. Keypunching and verification were based on 200 characters per questionnaire, 50 characters per address correction, and a rate of 7,000 characters per hour. Based on these assumptions and keypunch or clerical services at $6.25 per hour, processing costs are estimated at 56 cents per returned questionnaire. Processing of address corrections from postal returns is estimated at 9 cents for 30 percent of the followup system population.

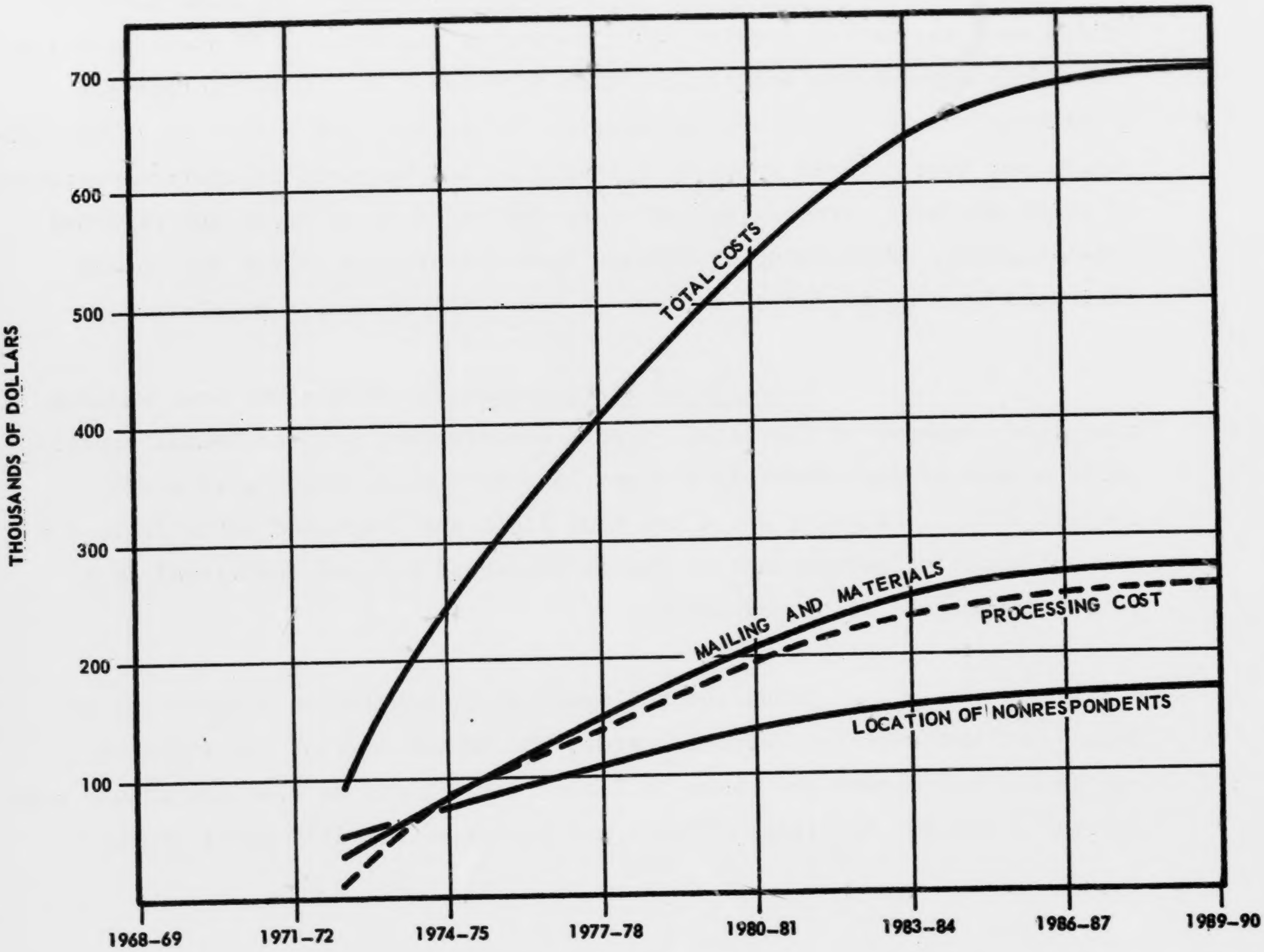
Location of Nonrespondents - With a one hour maximum time limit imposed on the Retail Credit Association, project TALENT experienced a cost of approximately $10 per located nonrespondent with a 72% success ratio. Allowing for a two hour limit and increased labor rates the cost of locating nonrespondents was estimated at $25 per individual in a 10 percent sample of nonrespondents.

Using the followup system population projections of Table II-5 and expected response rates, the annual cost of the followup system was calculated and shown in Table IV-6. There we find the annual cost starts at $93,000 in fiscal 1972-73 and increases steadily until fiscal

TABLE IV-6

FOLLOW-UP COSTS IN THOUSANDS OF DOLLARS

Year	Material & Mailing	Locating Non-Respondents	Processing Returned Documents	Total Cost
1972	35	50	8	93
1973	62	62	54	178
1974	87	75	78	240
1977	153	110	142	405
1980	209	137	195	541
1983	251	157	234	642
1986	268	165	251	684
1989	272	167	257	696



1989-90, where it starts stabilizing at $696,000. This assumes that a sample is drawn at the end of fiscal 1971-72 and questionnaires are mailed in fiscal 1972-73. As noted before, these figures do not include computer processing, administrative, or technical staffing costs. These latter items which include questionnaire design, analysis and reporting are included in the estimated staffing and processing costs of the overall system.

3. Central Processing Costs

Ideally, the estimates of central processing costs are made with detailed knowledge of the processing programs. Since the detail of the system and the programs is known only in general terms, these costs are estimated by relying upon the experience of the contractor for similar projects, and by the magnitude of the data handling and storage tasks.

To obtain an estimate of the amount of computer processing required for the proposed Student Flow Information System, an analysis was made of a system developed by the contractor covering some 365,000 individuals in 33 state operated institutions. The data record on an individual varies from 200 to 535 characters with all processing being accomplished on a medium scale computer. The processing averages 46 hours a week and includes:

- A weekly updating cycle of 3-1/4 hours
- Administrative reporting - 30 hours/month
- Research and statistical reporting process - 18 hours/week
- Special request reports - 15 hours/week
- Inquiry processing - 12 hours/month

The Student Flow Information System processing will be quite similar although the system will cover more than six times as many individuals initially. From the population data (Figure IV-1) the number of students or individuals included in the computer system after the first year of operation is 2.5 million. This will be increased by 500,000 records during the second year so that the total records in the system number 3 million at the end of that year. The average number of characters in a data record for this combined population will be approximately 400 characters. Thus, the data file is some 12×10^8 characters. Based on an average character storage capacity of 16×10^6 characters per reel of magnetic tape (800 b.p.i.), 75 reels will be required at the end of the second year for record storage.

Recognizing that the Student Flow Information System is quite large and that the updating process is quarterly, it appears reasonable to assume that processing operations for the Student Flow Information System will be between 40 and 60 hours per week when the system is fully operational.

The size of the processing requirement implies a minimum of a dedicated medium scale computer. This also implies that the processing could be accomplished on a large-scale computer which was devoted to the processing of a number of systems one of which is the proposed system. Either alternative would result in the same approximate processing costs. Thus, for the purpose of obtaining a cost estimate, the first alternative will be used.

A medium-scale computer when equipped with printer, card-read/punch, console and other normal devices, rents for about $25,000 per month or $300,000 annually. The staff to operate the computer on a two shift basis should include four operators (3.50/hour), one keypunch operator (2.50/hour), two programmers ($650/month), one librarian input-control clerk ($500/month) and one supervisor ($800/month). At the above rates and adding 35% for employer costs, the staffing cost approximates $100,000/year. The computer center and its staff will require approximately 3,000 square feet of prime space which rents for $1.00 per square foot per month. Thus, the total annual cost for space will be approximately $36,000. In summary the central processing cost will be approximately $436,000 annually - Figure IV-3.

If one compares the central processing cost with that of contracting with a service bureau to perform the computer processing, an indication of the accuracy of the central processing cost can be obtained. The 40 to 60 hours per week required for the processing of the proposed system and an approximate hourly rate of $200 per hour for a medium scale computer yield a total annual cost of $416,000 to $624,000. These figures are quite close to those shown in Figure IV-3.

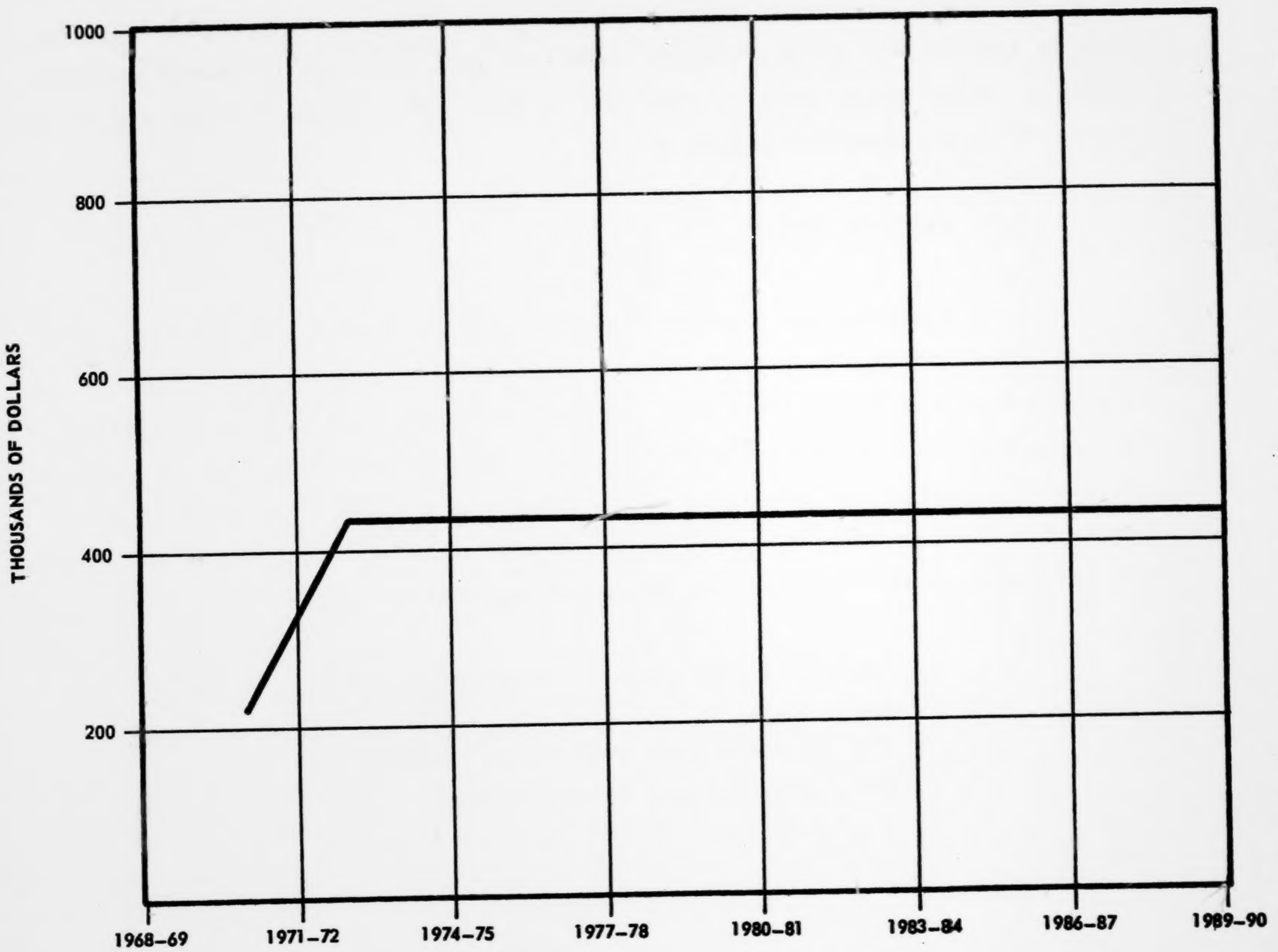
C. STAFFING COSTS

The proposed Student Flow Information System will gather and report upon information obtained from both public and private secondary and higher educational institutions. Because of such broad responsibilities, it is necessary that a permanent project staff be assigned the task of coordinating the design and operation of the proposed system. The project staff would have the overall responsibility of directing and monitoring the success of the system and in particular would be responsible for:

- Design of the computer processing system
- Development of questionnaires for use in educational and followup data collection systems
- Data analysis and interpretation

FIGURE IV-3

CENTRAL PROCESSING COST



- Preparation of reports
- Budgeting for the project
- Coordination with educational institutions

While the design of the computer processing system should be a contracted service, the project director and key technical assistants would be involved with the contractor from the beginning of the design phase (Fiscal 1969-70).

In addition to the normal secretarial help, this staff will require clerical assistance for mailing, filing, typing and similar activities. The clerical assistance requirement will increase from year to year as the volume of this effort increases.

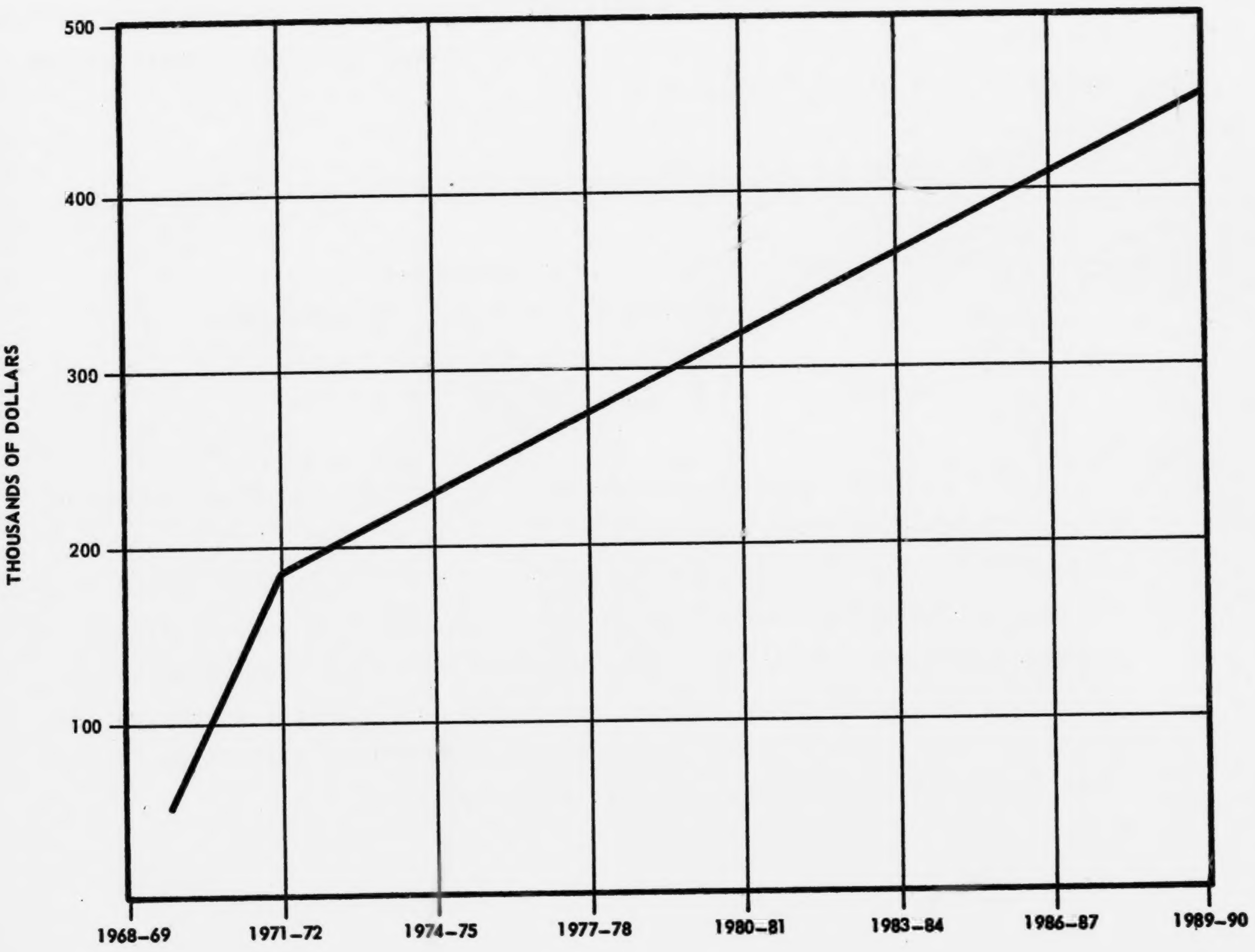
Costs for the staff have been determined as follows:

Positions: Director - $25,000/Year
Technical Assistant - $20,000/Year
Clerical - $7,200/Year
Overhead: Assumed to be 30%

Assuming that the project will be partially staffed (Director, three technical assistants and secretary) in fiscal 1969-70, the staffing costs will be approximately $50,000 in 1969-70 and $125,000 in fiscal 1970-71. Adding six clerks when the system becomes operational in fiscal 1971-72 increases the staffing costs to approximately $183,000. For purposes of estimating future staffing costs, it is assumed that salary increases and additional staff will add $15,000 each fiscal year from 1972-73 to 1989-90. These estimates are shown graphically in Figure IV-4.

FIGURE IV-4

PROJECT STAFFING COSTS



D. COST SUMMARY

The design, programming, operation and staffing costs are summarized in Table IV-8 and Figure IV-5 to provide a total cost for the proposed Student Flow Information System for each fiscal year. The estimated system cost is presented as a total cost obtained from the estimates of previous sections and a range which reflects the reasonableness of the estimates.

One of the most significant cost estimates is that of data collection. Accompanying this dollar significance is the vulnerability of some of the logic and projections of data upon which the cost estimate is based. An example is the selection of incremental characters which was made from the contractors experience and discussion with representatives of approximately ten institutions.

Using the assumption that the number of incremental characters in the calculations needs to be increased by 100%, the data collection costs must essentially be doubled. The total costs of operating the system each year would be increased between 40 and 50 percent. Though a doubling of the incremental characters to be collected has not been indicated by previous findings, a total cost based on this assumption indicates that the maximum total cost will not be more than 1-1/2 times the estimated cost presented in Table IV-8.

On the other hand, if the incremental data collected is halved, it could be expected that the total systems cost would be reduced by approximately 25 percent, thus indicating a minimum total cost of approximately 75 percent of the total estimated cost.

TABLE IV-7

COST SUMMARY
ESTIMATED TOTAL SYSTEM COST BY YEAR
(THOUSANDS OF DOLLARS)

Fiscal Year	Detail Design and Coordination	Programming and Implementation	Incremental Data Collection	Computer Facility	Project Staff	Followup	Estimated Total Cost		
							Minimum	Estimate	Maximum
1969-70	100	-	-	-	50(1)	-	-	150(2)	-
1970-71	50	100	400(3)	220(4)	125	-	-	895(5)	-
1971-72	-	50	611	436	183	-	975	1280	1891
1972-73	-	-	632	436	198	93	1043	1359	1991
1973-74	-	-	653	436	213	178	1154	1480	2133
1974-75	-	-	676	436	228	240	1242	1580	2256
1977-78	-	-	734	436	273	405	1481	1848	2582
1980-81	-	-	788	436	318	541	1689	2083	2871
1983-84	-	-	830	436	363	642	1856	2271	3101
1986-87	-	-	830	436	408	684	1943	2358	3188
1988-89	-	-	830	436	453	696	2000	2415	3245

---

(1) At least two full-time positions together with applicable clerical and other support should be provided within the Coordinating Council for Higher Education staff organization to superintend the system design work. These appointments should be provided whether the basic design work is accomplished with contractor support with technical teams from institutions or by some combination of resources.

(2) Start time approximately 1 July 1969.

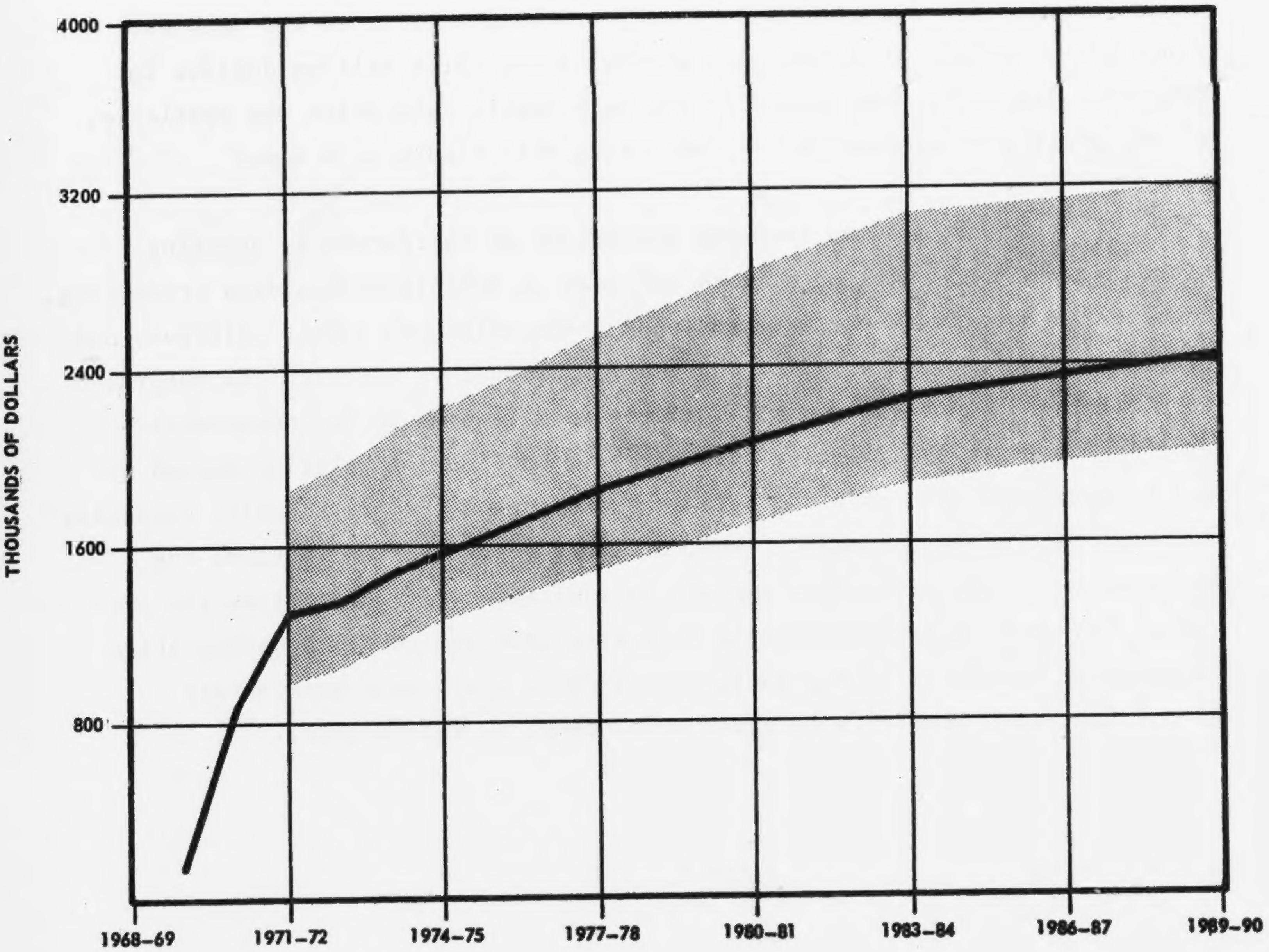
(3) Institutional costs associated with preparation of conversion programs, redesign of admissions and registration forms, and related activities.

(4) Includes cost of computer center, staff, supplies and equipment rental for six (6) months.

(5) Costs shown on this line and below are subject to revision during detailed design. Further, they will be incurred only if system implementation is approved in 1970-71.

FIGURE IV-5

TOTAL SYSTEMS COST BY YEAR



An attempt was made to determine the magnitude of the continuing investment which is being made by the educational institutions in administrative data processing. It is felt that this figure could provide an important insight into the extent which the educational institutions are relying upon data processing techniques to automate the information handling tasks of the institutions. One of the basic planks in the argument for the design of the proposed system is that the educational world is "going our way". If measures of the amount being spent were known, an estimate of the relative magnitude of the additional cost to be incurred by adding the proposed system to those being developed could be obtained. Unfortunately the usefulness of this information is exceeded by the difficulty of obtaining it. A measure of the annual cost of the administrative data processing activities was obtained from the University of California; approximately three million dollars in 1967-68. Since this was virtually the only usable data which was available, a conservative extrapolation was made using this figure as a base.

If one assumes that the University of California is spending approximately three million dollars per year in administrative data processing, one can conservatively estimate that the state colleges, junior colleges, and secondary schools will spend at least a like amount to maintain the information concerning their students. This estimate is seen to be conservative when one compares the relative size of the educational population served by the universities with that which is served by the secondary schools, community colleges, and state colleges. Thus, a figure of 12 million (4 times the expenditure by the university system) is indicated. If one applies the projected increase in expenditures in this area (12% per year), a conservative estimate of the level of expenditure in 1970-71 would be approximately 15 million; the actual figure would undoubtedly be much higher.

IV, D, Cost Summary (cont.)

By comparing this estimate with the total cost of the proposed system in fiscal 1971-72, one can see that approximately 8% would have to be added to the planned expenditures to obtain the benefits to be derived from the proposed Student Flow Information System.

V. SUMMARY

In previous chapters the reasons, rationale, design and cost of a Student Flow Information System have been developed. The major assumptions set forth were those of the increasing data processing capability of the various groups within the educational system and their ability to furnish a major portion of the data required by the proposed system. Arguments have been presented which indicate that the direction being taken by the educational institutions is one toward increasing use of data processing equipment. It is evident that developments are taking place within each educational group, but it is felt that the coordination between groups is slight.

A. COORDINATION

The contractor has met and discussed the problems of coordination with representatives of the University of California and other institutions. The task within the University of California is one of designing and developing common data systems for the University. They are directly engaged in coordinating the requirements of the various campuses of the University into one common information system effort--an effort which provides a common basis for the requirements of the university system as a whole and which is sufficiently flexible to permit the varying needs of each individual campus to be met. The representatives of the University of California, as well as many other individuals from the junior colleges and state colleges, have expressed on several occasions the need for coordination between educational groups. The desire for coordination is usually expressed concurrently with a realization that the proposed Student Flow Information System, being a system which requires data from all educational groups, can best provide the coordination desired.

Since the proposed system will require that data be furnished in common formats and coding structures, it will provide a tremendous impetus to establish uniform coding of data. Capitalizing on the requirements of the proposed system will provide the most effective and economical method of establishing common coding structures and paving the way for future information interchanges. With a small degree of foresight, one can readily comprehend the future importance of the type of information which will be collected within the institutions. The proposed system is but one of many requirements which will be established locally, within the state, and nationally. The commonality which was found between the information collected by each campus for its purpose and that information required by the proposed system will be repeated many times over.

Achieving the degree of coordination required will undoubtedly require substantial effort in the design phases of the proposed system. This effort has been recognized and included in the cost estimates. It has also been recognized that the additional burden of collecting information, which is not common to the institutional information system and the proposed system, is a burden which should be supported by the budget of the proposed system. A major portion of the estimated cost is thus derived from the reimbursement due the institutions for their efforts in collecting incremental data. Reimbursement is a new and revolutionary concept in the educational world. Currently, the state and federal governments require that certain information be prepared and forwarded to various agencies. The institutions have not been compensated for the efforts required in compiling such information. An example is the recent requirement for information concerning the ethnic origin of the students attending institutions of higher education throughout the country which is required by the U.S. Office of Education. It should be recognized that the added incentive from the reimbursements will have a positive effect upon the coordination effort. In the past, coordination has been argued on its merit alone. Never before, to our knowledge, has it been suggested that a small portion of the cost of such an effort would be financed from outside the institutions.

It has been the contractor's experience that most educational groups are in support of a coordinated effort toward the development of information systems. It is clear that the proposed system provides a great opportunity to enhance the coordination among the groups. Thus, it should be recognized that one of the perhaps most important benefits of the proposed system will be the practicality and ease of obtaining and combining information from the institutions within the educational system.

B. ORGANIZATIONAL IMPLICATIONS

When a system as large as the proposed Student Flow Information System is contemplated, there are a number of considerations which must be discussed. Two of these considerations have not been touched upon--organizational implications and confidentiality.

The success of a project of this magnitude can be influenced considerably by the organizational responsibility established for the project. The assignment of responsibility is significant since there is no one directing agency for education within the executive (or operating) areas of state government. The Department of Education is a coordinating agency for elementary and secondary education, and the Coordinating Council for Higher Education is the coordinating body for public higher institutions. Private educational facilities are private enterprises.

While the Joint Legislative Committee on Higher Education will supply the initiative to undertake this project, the operating responsibility cannot, with any logic, be assigned to any organizational entity within the Legislature. The Project Director and his staff should not be established as a separate organizational entity. Since the requirement for the system is primarily from the Legislature, the Director must be responsible to that body.

Thus, it is recommended the Director and his staff be assigned to the Coordinating Council for Higher Education, the advisory organization to the Legislature on matters relating to higher education.

The computing operations can and should be considered separately from the Project Director and his staff. None of the educational organizations as of this date have the computing capacity required for the Student Flow Information System. In general, these organizations are limited in capability to that required for their internal operation. Therefore, the assignment of the computer processing to any of these institutions must be accompanied by the appropriate resources to acquire staff and hardware. It is doubtful however that the operation should be directly related to any one educational institution.

The computer processing could be contracted through a service bureau. While the rates at service bureaus vary according to their particular charging practices, the cost of this alternative, as developed in Chapter IV, should approximate $416,000 to $623,000 per year. As was demonstrated in Chapter IV, this cost is comparable to the cost of operational staff, facility, and computer, and this alternative has the advantage of placing the complex operation in an experienced organization. Thus, the use of a service bureau appears to be an attractive alternative.

C. CONFIDENTIALITY

As with other systems where large amounts of data are related to individuals, there will be opposition to the establishment of the Student Flow Information System from those who are concerned about the privacy of the individual and the malicious use of information. Recognizing and agreeing with this concern leads to the investigation of methods of safeguarding the information and protecting the individual's right to privacy.

The proposed system depends heavily upon the ability to match students over a period of time. This matching can only be performed by relying upon information which identifies individuals. Thus, it becomes necessary to collect individual identity information and mandatory that this information be safeguarded. It is recommended that the personal identity information collected by the proposed system be protected by legislation which would restrict its use to the purposes for which it was originally collected. In particular, it is felt that certain groups (internal revenue, state franchise tax board, etc.) should be explicitly forbidden from using the information.

If any information from the proposed system is to be used in research projects, it should only be made available to researchers as data which cannot be related to individual persons. This can be done by eliminating all personal identify information prior to the use of the data base by any one other than that group which is responsible for the system. Similar projects such as TALENT and SCOPE have used this policy quite effectively to maintain the individual's privacy and to prevent the misuse of information.

It would be worthwhile to actively publicize the objectives of the proposed system and the safeguards which are an integral part of the system prior to its implementation. The experience of TALENT and SCOPE indicate that the most serious objections to systems of this type are the result of a misunderstanding of the goals of the system and an apparent lack of concern, on the part of those who advocate the system, for the right to privacy. It is easily seen that an ounce of prevention in publicizing widely the goals and safeguards which are an integral part of the proposed system could yield large dividends in cooperation and support.

D. PERFORMANCE

It is felt that design, programming, and operation of the Student Flow Information System could be best accomplished by relying upon the project director and his staff to direct the efforts of system design, programming and operation. Since an extensive staff of analysts and programmers has not been provided under the project director, it is reasonable to expect that the system design, programming, and possibly the operation will be performed under a contract with firms having capabilities in these areas. The coordination required in the systems design phases will require an unbiased effort--one which clearly cannot be performed by any one educational group. Thus, it is reasonable to expect that the project director and his staff perform as contract administrators who provide the impetus and direction of the effort but who do not actually perform the work.

E. TIMING

In previous chapters, the current and future efforts within each educational group have been presented and the need for a coordination of these efforts has been demonstrated. Since many of these efforts, particularly in the university system, have already begun, the timing considerations become very important. It will undoubtedly be less costly and difficult to integrate the requirements of the Student Flow Information System into the information systems of the educational groups prior to that time when the latter systems are fully operational. The best time to combine the data collection tasks will be during the design phases of both the proposed system and the campus information systems. Thus, it becomes necessary to look toward designing the proposed system at the earliest possible time. The time schedules presented in previous chapters represent a realistic estimate for the tasks which must be accomplished. Included in these estimates are time allowances for legislative approval, systems design, programming and implementation of the system.

V, E, Timing (cont.)

The Student Flow Information System offers the best potential solution to the problem of providing better information for the members of the California Legislature. The costs associated with the design, implementation and operation of such a system must be considered in comparison to the benefits that may accrue to the state, the educational system, and to the students themselves. It has been clearly demonstrated that similar efforts are currently in progress at both the secondary and higher educational institutions. If the coordination of the proposed system and those of the educational groups is to be effective, it is necessary to begin the detailed design as soon as possible.

REFERENCES

1. California's Needs for Additional Centers of Public Higher Education, State of California, Coordinating Council for Higher Education, Number 1014, December 1964.

2. California Population 1967, State of California, Department of Finance, October 1967.

3. Estimated and Projected Population of California 1960-2000, State of California, Department of Finance, June 1968.

4. Information Service Guide, The California State Colleges, Chancellor's Office, Sacramento, California.

5. Predicting Development of Young Adults, Cooley, W. W., and Lohnes, P. R., Project TALENT Office, American Institute for Research, Palo Alto, California, 1968.

6. SCOPE, Four-State Profile, Grade Nine 1966, California, Illinois, Massachusetts, North Carolina, The Center for Research and Development in Higher Education, University of California, Berkeley, California, December 1966.

7. "The Problem of Non-Response in Sample Surveys," Hansen, M. H., and Jurwitz, W. N., Journal of the American Statistical Association, 41 (1946), 512-529.

8. Total and Full-Time Enrollment, California Institutions of Higher Education, Fall 1967, State of California, Department of Finance, March 1968.

Preliminary Information Elements

Appendix A

PERSONAL IDENTITY INFORMATION

Name

Maiden Name

Address - Permanent
Current

Social Security Number

Selective Service Number

Sex

Ethnic Origin

Date of Birth

State or Country of Birth

Country of Citizenship

PARENTAL FAMILY BACKGROUND

- Parents
  - Living or Deceased
  - Marital Status

PRESENT FATHER OR MALE GUARDIAN

- Name
- Address
- Relationship
- Occupation - Type
- Income Range
- Highest Grade Completed
- Degree Level
- State or Country of Birth
- Year of Birth
- Country of Citizenship
- Bilingual

PRESENT MOTHER OR FEMALE GUARDIAN

- Name
- Address
- Relationship
- Occupation - Type
- Income Range
- Highest Grade Completed
- Degree Level
- State or Country of Birth
- Year of Birth
- Country of Citizenship
- Bilingual

BROTHERS AND SISTERS

- Number and Age of Each
- Number Entered College
- Number Graduated from College
- Other Legal Dependents of Parents

Appendix A

MARITAL FAMILY BACKGROUND

- Marital Status
- Spouse - Employed
    - Full - Part Time
    - Annual Income
    - Student
        - Full - Part Time
    - Highest Grade Level Attained
- Children
    - Date of Birth of Each
- Number of Other Legal Dependents

STUDENT BACKGROUND EDUCATIONAL DATA

- Last Elementary School Attended
    - What State
    - How Many Years
- Other High Schools Attended
    - What State
    - Number of Years in Each
- Other Colleges Attended
    - Name
    - Location
    - Number of Years
    - Degree
        - Level
        - Year Awarded
    - Grade Point Average
- Last High School Attended
    - What State
    - What County (California only)
    - How Many Years
    - Graduated?
    - Year of Graduation
    - Size of Graduating Class
    - Class Standing (percentile)
    - Test Results
        - Achievement
        - Aptitude
        - Other
- Vocational Schools Completed
    - Type
    - Year of Completion

Appendix A

EDUCATIONAL OBJECTIVES

- Level of Aspiration
  - Leave High School
  - High School Certificate
  - High School Diploma
  - Junior College Degree
  - Bachelor's Degree
  - Master's Degree
  - Doctor's Degree
  - Specialized Degree
- Vocational Aspiration
  - (Categorize)

CURRENT EDUCATIONAL DATA

- School Attending
  - Name
  - Location
  - Full/Part Time
  - Grade Level
  - Grade Point Average
  - Class Standing (percentile)
  - Major Field or Curriculum

STUDENT ECONOMIC DATA

- Employed
  - During School
    - Type Employment
    - School Related
    - Non-Related
  - Average Hours/Week
  - Weekly Earnings
- Vacation Time
  - Full/Part Time
  - Vacation Earnings
- Other Financial Resources (annual dollars)
  - Parental Support
  - Scholarships
  - Fellowships
  - G.I. Bill
  - Educational Loans
  - Company Reimbursement
  - Other
- Educational and Other Costs
  - Tuition and/or Fees
  - Books and Supplies
  - Room and Board
  - Travel
  - Other Related Expenses

Appendix A

FOLLOWUP DATA

- First Time for Each Exit
  - Reason for Leaving School
    - Completed Degree/Diploma
    - Health
    - Drafted
    - Financial
    - Required at Home
    - Grades
    - Disciplinary
    - Dissatisfied with School
    - Marriage
    - Other
- Major Field/Curriculum
- Currently Attending School
  - Name
  - Location
  - Full/Part Time
  - Grade Level
  - Grade Point Average
- Vocational Schools Attended
  - Type
  - Year of Completion
- Each Time Contacted
  - Schools Attended Since Last Contact
    - Name
    - Location
    - Number Semesters
    - Full/Part Time
    - Level Completed
      - None
      - High School Certificate
      - High School Diploma
      - Junior College
      - Bachelor's Degree
      - Master's Degree
      - Doctor's Degree
      - Specialized Degree
- Occupational History Since Last Contact
  - For Each Job Held During Period
    - Occupation Type
    - Annual Income
    - Number Years on Job
  - Present Employment
    - Full/Part Time
    - Occupation Type
    - Annual Income

Personnel Interviewed

Appendix B

Director of Admissions and
Registrar, University of California at Davis

Chancellor, California Community Colleges

Director of Admissions and
Registrar, Sacramento State College

Registrar, Sierra Jr. College

Assistant Superintendent, Special Services
and Programs Division, San Juan Unified
School District

Assistant Superintendent
Data Processing Manager,
Los Rios Jr. College District

Director of Data Processing,
Los Angeles City Schools

Director of Admissions
Registrar
University of California at Los Angeles

Director, Division of Research,
Chancellor's Office, California State Colleges

Representative of Institutional Studies,
Long Beach State College

Assistant Dean of Admissions,
University of the Pacific

Data Elements for Students While Enrolled in California Educational Institutions

Appendix C

A. Student Identity Data

1. Name
2. Social Security Number
3. Sex
4. Date and Place of Birth
5. Ethnic Origin (to include foreign student)
6. Permanent Address

B. Parental Family Data

1. Name of a Parent or Guardian
2. Relationship
3. Occupation
4. Educational Attainment
5. Permanent Address

C. Marital Family Data

1. Marital Status (single or married)
2. Number of Dependents - Age of Each
3. Student Employment - Number of Hours per Week

D. Student Background Educational Data

1. Last High School Attended
   Name
   State (and County if California)
2. Year of Graduation from High School
3. High School Grades (or other index of achievement)
4. Other Colleges Attended - Name and Location
5. Degrees Obtained and Year Awarded
6. SAT/ACT Test Scores

E. Current Educational Data

1. School Attending - Name and Location
2. Grade Level (number of units completed)
3. Major Field or Curriculum
4. Full or Part Time (number of units enrolled in)

F. Plans or Objectives

1. Educational Objective (degree level)
2. Vocational Aspirations

Final Data Elements

Appendix D

Admissions:	No. Characters
1. Student Identity	
Name	30
Social Security Number	9
Sex	1
Date of Birth	6
Country of Birth/State if USA/County if California	6
Ethnic Origin (to include foreign student)	2
2. Parental Family Data	
Male Parent or Guardian	
Relationship	1
Occupation	2
Educational Attainment	1
Female Parent or Guardian	
Relationship	1
Occupation	2
Educational Attainment	1
3. Student's Educational Background	
Last High School Attended	6
Month, Year of Graduation from High School	4
High School Grades (point average or other achievement index)	10
Last College Attended	6
Grade Point Average (last college attended)	3
Degrees Obtained and Month/Year Awarded (maximum 4 degrees)	
SAT/ACT/Other Test Scores	18
4. Current Educational Data	
School Enrolling In	6

	No. Characters
5. Plans or Objectives	
Educational Objectives (degree level)	2
Vocational Aspirations	2
TOTAL	139

Registration:

1. Student Identity	
Name (last, first, middle)	30
Social Security Number	9
Permanent Address	41
2. Parental Family Data	
Name of Parent or Guardian or Person who will know your location	30
Address of Above	41
3. Marital Family Data	
Marital Status (single or married)	1
Number of Legal Dependents	2
Age of Each Dependent	2/Dependent (assume three)
Student Employment - No. Hours per Week during school semester	2
4. Current Educational Data	
School Attending	6
Grade Level (number of units completed)	3
Major Field or Curriculum	4
Full or Part Time (number of units enrolled in)	2
Grade Point Average to Date	3
TOTAL	180

Appendix D

Graduation:	No. Characters
Name (last, first, middle)	30
Social Security Number	9
Date of Graduation	4
Degree Awarded	1
Major Field or Curriculum	4
Grade Point Average	3
School of Graduation	6
TOTAL	57

Followup Data

Appendix E

1. Identity and Location Data

Name	30
Social Security Number	9
Date of Birth	6
Address	41
Name of Parent, Guardian, or Other Person who will know your location	30
Address of Above Individual	41

2. Marital Family Data

Marital Status	1
Number of Dependents	2
Age of Each Dependent	2/Dependent

3. Current Educational Data

Whether or Not Currently Attending School	1
Name and Location of School Attending	6
Full or Part Time Student (No. units enrolled in)	2
Grade Level	1
Major Field or Curriculum	4
Educational Objective (degree level)	2

4. Educational History Data

Reason Left California Educational Institution (first contact only)	2
Other Schools Attended Since Last Contact	6/School
Number of Units Completed Each School	3/School
Degrees Awarded and Year Obtained	5/Degree

5. Vocational Training Data

Type Vocational Training Completed	2
Month and Year Completed	4

Appendix E

6. Occupational and Income Data

Whether or Not Currently Employed	1
Current Occupation	2
Full/Part Time Employment	1
Number of Years Present Occupation	2
Annual Income	1
Vocational Objective	2
Spouse Employed	1
Spouse's Occupation	2
Spouse's Annual Income	1
TOTAL	217*

---

*Assumes three dependents, one school and one degree since last contact.

Incremental Characters

Appendix F

Class of Data	High School	Jr. College	State College	University
ADMISSIONS				
Student Identity				
Social Security	9			
Ethnic Origin		2	2	2
Parental Family				
Relationship		1	1	1
Occupation		2	2	2
Educational Attainment		1	1	1
Student Educational Background				
Year of Graduation/High School		2	2	2
High School Grades		10	10	10
SAT/ACT/Other		18	18	18
Current Educational Plans or Objectives				
Educational Objectives	2	2	2	2
Vocational Aspirations	2	2	2	2
Total	13	40	40	40
REGISTRATION				
Student Identity				
Social Security	9			
Parental Family Data				
Name of Parent/Guardian or Person knowing your location		30	30	30
Address of Above		41	41	41
Marital Family Data				
Number of Legal Dependents	2	2	2	2
Age of Each Dependent	2/Dependent	2/Dep.	2/Dep.	2/Dep.
Student Employment	2	2	2	2
Current Educational Data				
Grade Point Average		3	3	3
Total	19	84	84	84
GRADUATION				
Total	0	0	0	0

NOTE: Assumes three dependents.

AEROJET
GENERAL TIRE
GENERAL

AEROJET
GENERAL TIRE
GENERAL
®

AEROJET-GENERAL CORPORATION
COMPUTING SCIENCES
P. O. BOX 15847
SACRAMENTO, CALIFORNIA 95813
TELEPHONE 916/355-2540